

**WHEN SYSTEM DEMANDS MEET SITE REALITIES
IN HIGH-STAKES LITERACY TESTING:
A TASMANIAN EDUCATION CASE STUDY**

by

Lauren Margaret Johnson

Bachelor of Arts (Hons.), Bachelor of Teaching (Hons.)

Submitted in fulfilment of the
requirements for the Degree of
Doctor of Philosophy

University of Tasmania

July, 2011

DECLARATION OF ORIGINALITY

This thesis contains no material which has been accepted for a degree or diploma by this University or any other institution, except by way of background information and duly acknowledged in the thesis, and, to the best of my knowledge and belief, it contains no material previously published or written by another person except where due acknowledgement is made in the text of the thesis, and nor does the thesis contain any material that infringes copyright.

Lauren Johnson

AUTHORITY OF ACCESS

This thesis may be made available for loan and limited copying and communication in accordance with the *Copyright Act 1968*.

Lauren Johnson

STATEMENT OF ETHICAL CONDUCT

The research associated with this thesis abides by the international and Australian codes on human and animal experimentation, the guidelines by the Australian Government's Office of the Gene Technology Regulator and the rulings of the Safety, Ethics and Institutional Biosafety Committees of the University.

Lauren Johnson

ABSTRACT

Literacy testing is regarded as high and ever-increasing stakes, in Australia and beyond. The value and validity of testing and the tests themselves, uses made of test data and pressures to improve scores represent conflict for teachers. This topic has immediate significance for Tasmanian schools engaged in testing programmes, and for the wider education systems in Australia and internationally. This research examines these issues in this contested field with a focus on the lived experience of those most closely involved.

This thesis explores the discursive tensions and conflicts within secondary teachers' experiences of standardised literacy testing in Tasmania. The research was conducted through a case study of one secondary-level State school in Tasmania involved with government-mandated standardised literacy tests. Testing sessions were observed and interviews conducted during one round of testing. Two research questions focus the analysis on teacher perceptions and test administration in the classroom site.

Through a combined constructivist grounded theory (Charmaz, 2000, 2001, 2005, 2006; Charmaz & Mitchell, 2002) and discourse analysis (Carabine, 2001; Fairclough, 2001; Taylor, 2001a, 2001b) approach, emergent patterns and themes are examined. This analytic approach enabled the identification of dominant discourses (Gee, 1998, 2003, 2005, 2011), discursive tension and conflict.

Of particular importance in this study are the various teacher perspectives on their roles as test administrators. Participants' words and voices are studied to examine the ways that teacher perspectives affect their administration of the tests, and how their perceptions of standardised testing can become transmitted to students through teacher language and behaviours.

The recognition of often-competing demands within the education system and school sites (Freebody & Wyatt-Smith, 2004) is highly relevant to this research, given that standardised literacy testing happens at the intersection where system demands meet site realities. This research notes the discursive tension and conflict resultant of system-site (dis)connectivity.

The research project contributes a critical understanding of standardised literacy test administration, necessary for deeper and more nuanced understanding of what is valued and devalued through such testing, and how school test actors respond to competing test demands. This research recognises the ways that system pressure

for educational testing uniformity affects teachers, students and the wider school setting. Discursive tensions have implications for test administration and school management, within and beyond the Australian education context. The research identifies a number of such implications as findings, and makes considered recommendations for future research, policy and test design, and professional practice.

ACKNOWLEDGEMENTS

I must first acknowledge and thank my three main supervisors: Dr. Claire Hiller, Prof. John Williamson and A/Prof. Geraldine Castleton. I very much appreciate Dr. Hiller's initial discussion with me of this topic to which I immediately felt strongly drawn. I also wish to express my gratitude to Dr. Yoshi Budd, for years of unofficial and tireless mentoring. I can't thank you enough. Thanks also to those who participated in this research, directly and indirectly, allowing me to see into your world and interpret what I saw.

Universes of thanks go to my mum, Dr. Karen Vincent, for all her help throughout the stages of writing this dissertation. I simply could not have completed this work without your help. Huge appreciation goes to Myron, whose IT support saved my drafts more times than either of us had hoped or wanted, and whose support in other ways saw me through to submitting. Great thanks go to all my family and friends for their support in various forms through my PhD studies, though some of you decided to make my completion difficult by getting married, having babies, moving away and tempting me with lingering coffee shop brunches and Sunday roasts.

Many thanks to my somewhat large nuclear and step-family(ies), particularly my parents and brother, who kept alive my passion for scholarship and research, and to my teaching colleagues and students, who provided me with daily reminders of the importance of teachers' work – though I freely admit to being thoroughly relieved to reach the end of my PhD time.

CONTENTS

Declaration of originality	i
Authority of access	ii
Statement of ethical conduct	iii
Abstract	iv
Acknowledgements	vi
Contents.....	vii
List of tables	x
List of figures	x
Acronyms	xi
Research participant acronyms	xi
Note on terms	xi
Chapter One: Introduction.....	1
Introduction.....	1
1.1 Study Focus and Purpose of the Research.....	10
1.2 Study Context.....	14
1.3 Analysis and Methods	15
1.4 Structure of Thesis.....	17
1.5 Summary	19
Chapter Two: Literature Review.....	21
Introduction.....	21
2.1 Literacy and Literacy Testing.....	22
2.1.1 Literacy in Australia	23
2.1.2 Trends in standardised literacy testing.....	24
2.2 The Classroom Site.....	61
2.2.1 Teachers and students in the classroom.	61
2.2.2 Educational change at the school site	69
2.2.3 Teachers' mediation of literacy assessment.....	76
2.3 Summary of Literature Review	80
Chapter Three: Methodology	83
Introduction.....	83
3.1 Research Questions	84
3.2 Selection of the Research Approach.....	87
3.2.1 Qualitative research	87
3.2.2 Case study	88
3.3 Theoretical Principles Underpinning the Research	89
3.3.1 Postmodernist and poststructuralist theory	89

3.3.2	Discourse.....	91
3.3.3	Subjectivity and researcher reflexivity.....	96
3.3.4	Problematising validity, reliability and credibility.....	99
3.4	Development Phase of the Research.....	101
3.4.1	Initial research approval: Ethics.....	103
3.4.2	Site selection	104
3.4.3	Sample.....	106
3.4.4	Information letters and consent forms.....	107
3.4.5	Participants.....	107
3.4.6	Development of research instruments.....	109
3.4.7	Approval of research instruments	109
3.5	Data Generation and Collection.....	110
3.5.1	Testing session transcript data and field notes.....	112
3.5.2	Questionnaire data.....	114
3.5.3	Interview data.....	117
3.6	Approaches to Data Analysis.....	121
3.6.1	Constructivist grounded theory.....	123
3.6.2	Discourse analysis.....	130
3.7	Structure of Analysis Chapters	136
Chapter Four: Coding Analysis of Data.....		138
Introduction.....		138
4.1	Details of Coding Analysis	140
4.2	Initial Coding	143
4.2.1	Initial coding: Participant interviews	144
4.2.2	Initial coding: Testing session transcripts	151
4.2.3	Summary of codes.....	156
4.3	Focused coding	156
4.3.1	Focused coding: Participant interviews.....	157
4.3.2	Focused coding: Testing session transcripts	164
4.3.3	Summary of categories.....	172
4.4	Axial coding.....	172
4.4.1	Theme 1: Maintenance of control, behaviour management, order.....	174
4.4.2	Theme 2: Utility of standardised literacy tests as educational/learning tools.	175
4.4.3	Theme 3: Pragmatics of standardised literacy tests.....	176
4.4.4	Theme 4: Judgements on student communities, standards, attitudes	177
4.4.5	Theme 5: Perceptions and behaviours of teaching.....	178
4.5	Summary of Coding Analysis.....	179
Chapter Five: Research Questions and Themes.....		182
Introduction.....		182
5.1	Research Question 1	187
5.1.1	Literacy diagnoses.....	188
5.1.2	School-wide literacy programme planning and test timeframe issues	193
5.1.3	Political use made of test data.....	197

5.1.4 Detail provided in test data	201
5.1.5 Influences upon test sessions	203
5.1.5 Summary of research question 1	207
5.2 Research Question 2	207
5.2.1 Clarification and affirmation of test rules, instructions and expectations	209
5.2.2 Responding to student questions and concerns of testing	217
5.2.3 Addressing student negativity towards testing	235
5.2.4 Summary of research question 2	242
5.3 Results	243
Chapter Six: Discourse Analysis	247
Introduction	247
6.1 Defining the Discourses	249
6.2 Dominant Discourses and Discursive Tensions	251
6.2.1 Discourse of Uniformity	251
6.2.2 Discourse of Reflexivity	254
6.2.3 Discursive tensions	256
6.3 Effects of the Discourses	261
6.4 Absences and Silences within the Discourses	265
6.5 Key Findings	267
Chapter Seven: Conclusion	281
Introduction	281
7.1 Research and Limitations	281
7.1.1 Overview of the research	281
7.1.2 Possible limitations	283
7.2 Key Findings	285
7.3 Recommendations	286
7.3.1 Future research	286
7.3.2 Policy and professional practice	288
7.4 Research Contribution	290
References	294
Appendix	321
Appendix A: Approval Documentation	323
Appendix B: Data Instruments	341
Appendix C: Data Extracts	350
Appendix D: List of Codes by Frequency	358

LIST OF TABLES

Table 3-1. Timeline of data collection.....	102
Table 3-2. Steps taken within each stage of data collection.....	103
Table 4-1. Interview transcripts: Codes and frequencies.....	145
Table 4-2. Interview transcripts: Most frequent codes.....	148
Table 4-3. Testing session transcripts: Codes and frequencies.....	152
Table 4-4. Testing session transcripts: Most frequent codes.....	154

LIST OF FIGURES

Figure 3-1. Codes 1-3, interview transcripts and testing session transcripts.....	125
Figure 3-2. Category 1 (interview and test session), with three codes.....	126
Figure 3-3. Theme 1 with categories.....	128
Figure 3-4. Carabine's (2001) adapted discourse analysis steps.....	133
Figure 3-5. Fairclough's (2001) adapted discourse analysis steps.....	135
Figure 4-1. Coding example.....	141
Figure 4-2. Interview transcripts: Codes and categories.....	158
Figure 4-3. Testing session transcripts: Codes and categories.....	165
Figure 4-4. Themes.....	173
Figure 4-5. Themes and categories.....	174
Figure 5-1. Themes.....	185

ACRONYMS

ACARA:	Australian Curriculum, Assessment and Reporting Authority
AEU:	Australian Education Union
DoE:	Department of Education
NAPLAN:	National Assessment Program – Literacy and Numeracy
OECD:	Organisation for Economic Co-operation and Development
PISA:	Programme for International Student Assessment
UK:	United Kingdom
US:	United States of America

RESEARCH PARTICIPANT ACRONYMS

A1, A2:	Teacher Aide 1, 2
AP:	Assistant Principal
P:	Principal (research participant)
TT:	Testing Teacher (research participant)
T1:	Teacher 1 (research participant)
T2:	Teacher 2 (research participant)
T3:	Teacher 3

NOTE ON TERMS

Secondary level: This term refers to year or grade levels 7 to 10 or 7 to 12 (Years 11 and 12 are also referred to in Tasmania more specifically as Senior Secondary). Students in Years 7 to 9 are usually aged between 11 and 15 years, dependent on such factors as birth month and repetition of school years.

State school: The term ‘State’ or ‘government’ school is often used in Australia to refer to a ‘public’, as distinct from a ‘private’, school. Australian State governments traditionally provide the bulk of the funding for ‘public’ schools, and Federal/Commonwealth governments in turn contribute the bulk of government funding provided to ‘private’ schools.

Chapter One: Introduction

Introduction

The idea of a crisis in literacy has been subject to considerable discussion in Australian literature and countries with similar education systems, especially the United Kingdom (UK) and United States of America (US). Debates regarding literacy standards, and the inextricably connected topic of testing and measuring such standards, are common and impassioned; perhaps unsurprisingly so, as education and educational policy on a broad scale are always political (Henig, 2009). Examining Australian literacy levels against international comparisons helps to contextualise the debate and explain how dominant thinking has shaped ideas about the crisis and education policy responses.

Literacy represents a tension-point for teachers and for those involved in literacy education outside the (school) site, according to a number of writers, because of this perceived crisis in and of literacy, the responsibility for which “educational systems and educators” (Green, Hodgins, & Luke, 1997, p. 8) are left to shoulder (Christensen, 2008; Comber & Kamler, 2004; Hodgetts, 2010; Kosar, 2005; Lingard, 2010; Woods, 2007). Among changes imposed on the classroom from the system level and beyond, moves towards educational standardisation reflect longer-term overseas trends, notably in the UK and the US (Berliner & Biddle, 1995; Caldwell, 2010; Hursh, 2008; Kohn, 2000a, 2000b; Koretz, 2008; Sacks, 2000; Wyatt-Smith & Klenowski, 2010). Such changes have affected teachers’ roles and what happens in the classroom with

students, and are not localised to these three countries (Williamson & Myhill, 2008).

The international education context within which Australia is located values comparison of achievement standards, which can highlight educational systems' (and actors') relative strengths and weaknesses. As such, Australia takes part in a number of international educational comparisons, including the OECD's 'Programme for International Student Assessment' (PISA), 'Trends in International Mathematics and Science Study' (TIMSS) and, from 2011, the 'Progress in International Reading Literacy Study' (PIRLS); in such international tests, Australian students overall rate highly (OECD, 2010; Thomson & Buckley, 2007), although across OECD countries, parents and those at the education system level nonetheless value further improvements (Butler & van Zanten, 2007). Within Australia, disparity in literacy levels is evident along student background, geographical location and sex, and across the states and territories, with Tasmania usually towards the bottom of the state/territory rankings (McGaw, 2007; Nguyen, 2010; OECD, 2009, 2010; Wyatt-Smith, 2008), and with great disparity even within classrooms (Ladwig, 2010).

In addition to taking part in international literacy evaluations, the Australian Federal government mandates student involvement in nationally run literacy achievement tests, with which tests this research is specifically concerned. The standardised literacy testing processes examined in this research have only recently become standardised across the Australian states and territories (Wyatt-Smith, 2008; Wyatt-Smith & Klenowski, 2010), when in

May 2008 the first nationally uniform literacy and numeracy testing, the National Assessment Program – Literacy and Numeracy (NAPLAN), was introduced. Prior to that date, “[d]espite being part of a national testing program, the tests for each state [were] developed within that state and there [was] no equating of tests at a statistical level across states” (Lake Corporate Consulting, 2006, p. 2). These early state-based tests were also only relatively recent initiatives, following a period of one to two decades of minimal systematic monitoring of Australian educational outcomes (Ladwig, 2010).

Through NAPLAN, identical tests are used for students in each of the year levels Three, Five, Seven and Nine. The tests combine multiple-choice and written responses; completed tests are computer- and manually-marked. NAPLAN “tests are promoted as a means of bridging perceived gaps in accountability, and providing the data through which the quality of education can be improved” (Willett & Gardiner, 2009, p. 2). As Research Director of the Australian Council for Educational Research’s Systemwide Testing Program, Freeman (2009) proposed expanding NAPLAN’s goals:

In the future, NAPLAN results may be used in evaluating the effectiveness of educational policies and programs; in identifying the need for targeted interventions for individual students and groups of students; and in developing improved measures of school performance and greater transparency in national reporting. ... The tests are constructed to assess knowledge, skills and understandings appropriate to each year level; to be interesting and engaging to students throughout Australia; and to challenge students at all levels of ability. (p. 13)

Freeman's (2009) comments above serve to summarise debate with NAPLAN and high-stakes testing in Australia concerning test capability and purpose distinct from idealised future uses. It is acknowledged by Freeman that these tests cannot test every element of literacy, and cannot in their current form be utilised to make detailed evaluations of student results or longer-term goals. For example, NAPLAN literacy tests, as with other standardised literacy tests, only test "aspects" (Lake Corporate Consulting, 2006, p. 1) of literacy; they measure "students' performances in literacy, including reading, writing, spelling, grammar and punctuation" (Freeman, p. 13). The tests can, however, provide quantitative measurements of students' performance on particular aspects of literacy, upon which action can be taken (Fremer, 2005).

At the time of data collection for this research, shortly before the introduction of NAPLAN testing, Tasmanian students (Year Seven and Nine, who took part in this research) sat three tests within the literacy grouping: reading/comprehension, writing/composition, and spelling. The recently-introduced NAPLAN tests are slightly different, again with three literacy tests but testing reading, writing, and language conventions, the latter including spelling, grammar and punctuation (ACARA, 2010a). Tests are held over the same days across the country, with uniform guidelines provided for test administrators and students. Test results are returned in two stages later in the school year of testing:

The first stage Summary Report is released in September prior to the distribution of reports to parents. This report shows results at each year level and domain by state and territory and nationally.

The second stage is the full National Report that includes detailed results by gender, Indigenous status, language background other than English status, parental occupation, parental education, and geolocation (metropolitan, provincial, remote and very remote) at each year level and for each domain of the test. This report, which relies on more detailed analysis, is provided at the end of the year. (ACARA, 2010b)

For the 2008-2010 NAPLAN tests, Tasmanian Year Seven and Nine students' literacy results were no more than two percentage points below the national average on any of the literacy tests, and the 2008 Year Nine Reading result in Tasmania was 0.1% above the national average (ACARA, 2010c). For the 2009-2010 literacy reporting, Tasmanian Year Seven and Nine literacy scores in all three tested areas showed no significant statistical difference from previous the 2008 or 2009 tests (ACARA, 2010c).

This profile can contextualise Tasmanian discussion of test results, taken publicly as comprehensive evidence of education achievements, by raising doubts about an assumed literacy crisis compared against other Australian states/territories; nonetheless, as already noted, Australian students' test results indicate differential and inequitable literacy outcomes across students' geographical location (such as rural or urban), sex, socioeconomic status and student background, and quality of teachers (OECD, 2009, 2010; Pegg & Panizzon, 2007). As such, the topic of literacy testing and student literacy standards remains contested as public debate continues and anxiety for those involved at the school site can become heightened at testing periods.

Proponents of educational standardisation and accountability variously claim to be attempting to redress imbalances caused by critical literacy, whole

language, and political correctness in education (Donnelly, 2004; Henderson, 2005), and that clarifying students' learning progress enables improvement (Cohen, 2010; Kosar, 2005; Lubienski, Gulosino, & Weitzel, 2009). Critics, meanwhile, argue that these educational trends of standardisation and increased accountability, also experienced in other human services fields (Comber & Nixon, 2009), serve to increase government, public and media control over what happens at the school site, constricting teachers' professional autonomy and not necessarily benefitting students' literacy outcomes (Beck, 2008; Berliner & Biddle, 1995; Davies & Bansel, 2007; Hill, 2005; Hursh, 2008; Kohn, 2000a, 2000b; Purpel, 1999; Sacks, 2000). Standardised literacy testing is an especially public source of contention within these broader educational trends and has come to represent 'high-stakes' politically in Australia and elsewhere, as a result of uses made of test data and misconceptions and debate about how such data are generated and should be used (Caldwell, 2010; Koretz, 2008; Lingard, 2010; Wyatt-Smith & Klenowski, 2010). Even Australia's literacy benchmarks, against which students' literacy levels are tested and compared, have been identified as cause for debate and criticism (Freebody, 1998; Gill, 1998; Knapp, 1998).

Teachers' direct involvement in test content and processes too has been the basis for debate in Australia. The organisation responsible for designing the NAPLAN tests, the Australian Curriculum, Assessment and Reporting Authority (ACARA), claims that teachers are consulted as to the content of the tests (Craig, 2010; Freeman, 2009). Australian teachers' reactions to the tests as well as to the tests' reporting on 'My School', however, have been

conflicted, and Australian teachers have threatened test boycotts (Hudson & Masanauskas, 2010; Jensen, 2010b). These debates and contradictions clearly show that this research project is located in a contested field with stressors and pressures from a range of sources.

Among these debates that surround high stakes, standardised literacy testing in Australia (and elsewhere) are those that place it in a broader political context. In recent decades, Australian public policy has become increasingly shaped by a pervasive market ideology (Davies & Bansel, 2007; Howard, 2003), punctuated in the education policy arena by research demonstrating the benefits of high literacy standards for a number of reasons, including economic, health, employment and personal development and wellbeing (Castleton, 2010; Dugdale & Clark, 2008). The relationship between educational outcomes and international economic competitiveness has particular resonance for Australian state and federal governments (S. Black, 2004), leading to the current nationwide mandated testing of students against set standards and benchmarks (Comber & Nixon, 2009). Linking, and at times tying, government funds to selected education programmes, in particular to test results, has led to a label of 'high stakes' being attached to such measures (Caldwell, 2010). This approach has been seen as necessary at national and state levels to achieve increased public accountability of schools in response to a range of dissatisfactions, and by extension to government policy efforts that locate education in broader public policy.

A key manifestation of market ideology in the education arena is school comparison and competition. There is a degree of ambivalence and tension

among education professionals towards educational accountability through high-stakes, standardised literacy testing and the public reporting of school test results (Wyatt-Smith, 1998; Wyatt-Smith & Klenowski, 2010). School comparison, competition and reporting of school results is attempted through the publication of Australian schools' NAPLAN ratings and comparative status on the Federal government's 'My School' website. Such moves are couched within business arguments that insist competition and pressure force under-performing schools to improve, and ineffective processes to be identified (Lubienski, Gulosino & Weitzel, 2009). The Independent Schools Council of Australia, for instance, publicly lauded these trends through acknowledgement and acceptance that public perceptions of independent schools' reputations, including academic standards, allow these schools to "survive or fail" (Daniels, 2005, p. 2). In contrast, such moves are also made against claims that comparison and competition contribute to labels of successes and failures that do not address the causes of differential educational outcomes (Merrett, 2006).

As such, the use made of such testing and test data directly affects those working at the (school) site. Doecke, Reynolds and Roberts (2002), while apparently not disputing government claims to want to improve teachers' status, identified that teachers' professionalism and autonomy are potentially undermined by externally-imposed tests of literacy levels, the publication of test results and decisions based on those results. Parr and Bellis (2006), in examining neoliberal and market discourses in education, pointed to an educational consumerism that influences perceptions of teacher efficacy:

teachers need to accede to consumer demand instead of their own professional expertise, to the expense of “English teachers’ professionalism” (p. 7).

In addition to the use made of tests, particularly heated debate surrounds issues related to the accuracy or validity of test results. Although Kosar (2005) claimed that this aspect of such testing garners little or no scholarly attention, and that the accuracy and validity of test results is evident, researchers are still divided as to whether tests measure what they are intended or purported to measure and whether the tests can in fact represent student literacy learning (Hursh, 2008; Kohn, 2000b; Koretz, 2008; Ohanian, 1999; Phelps, 2003). In this regard, validity can refer to content, concurrent and predictive validity and test reliability (Wilde, 2002).

This research is located in response to these education trends and debates, with particular reference to increasing standardisation in literacy teaching, learning, assessment, testing, monitoring and reporting. In Australia, this trend has appeared to come from both Federal and State governments. Since January 2005, the beginning of this research project, Australian Federal politics has appeared to move towards a model of increasing standardisation in education, not only in literacy standards and school league tables (‘My School’), but also with A-E reporting and a national curriculum.

Standardised literacy testing has become widely popular and relied-upon as the tool preferred by Federal and state governments to measure and compile, in quantitative and therefore comparable form, student literacy levels and standards. These tools are becoming so standardised that Tasmanian students’ performance on the three branches of the literacy tests can be

compared with the performance of their peers in other schools, school districts and states/territories. These tests are standardised in content and conditions: students in any given year level in the same school year sitting their literacy tests are tested with identical questions, tasks, directions, time provisions and materials. The issues of the uniformity of these tests, uses made of student test data, and involvement of various interest groups in student literacy standards and testing, have all generated considerable debate. Those interested in standardised literacy testing include parents, education unions, media, politicians and government, and business and industry, as well as classroom test actors. This term ‘test actors’ is used in this research to denote those directly involved in standardised literacy testing at the school site, such as students and teachers.

1.1 Study Focus and Purpose of the Research

This study analyses teachers’ perspectives on their roles and experiences as test administrators, and explores the discursive tensions and conflicts within secondary teachers’ experiences of standardised literacy testing processes and practices at the school site. This study examines the topic through the context of Tasmanian secondary-level State-school standardised literacy testing, within the frame of two related research questions.

The purpose of the research, refined through the literature review, is to address a significant vacuum within Australian school-based research on teachers’ perspectives, experiences and approaches regarding standardised literacy testing. The broad area of standardised literacy testing has been the

subject of much research and debate, including such aspects as the politics of testing, debates of equity and objectivity, and uses made of the tests. An initial literature survey revealed a lack of knowledge in schools research concerning teachers' perspectives, experiences and approaches regarding the tests. The literature utilised in this work was located through following references in relevant works as well as searching online databases, most particularly ERIC (Education Resources Information Center) and JSTOR Arts and Sciences.

The school-based, teacher-focused approach employed in this study was selected in order to produce research more grounded in school site realities, and findings more directly relevant to the school site of testing. As such, the two research questions employed in this research clarify the foci of the study. The dual research focus is teachers' perceptions and interpretations of the value and validity of standardised literacy testing, and action (teachers' test administration) and interaction throughout and regarding testing.

This study utilises constructivist grounded theory (Charmaz, 2000, 2001, 2005, 2006; Charmaz & Mitchell, 2002; Corbin & Strauss, 1990; Haig, 1995) and discourse analysis (Carabine, 2001; Fairclough, 2001; Taylor, 2001a, 2001b) approaches, as a result of which the focus and direction of the research is informed and refined by the data analysed. Following the selection of the initial area of research (standardised literacy testing), the specific aspects of standardised literacy testing to be examined were clarified only during the data generation, collection and analysis stages. Such an inductive research process is considered most appropriate and beneficial for the produced research work, as it allows the topic to be as general as possible within the frame of

standardised literacy testing to the point where the data begin to shape the specific direction of the study.

The data began to focus on the administration of the tests quite early in the research process. This was perhaps unsurprising, as the study participants were teachers (the participating principal also retained a teaching role), and teachers' roles through standardised literacy testing processes are mainly those of test administrators. Test administration, then, is the area of standardised literacy testing with which the selected participants were most familiar, though the principal also experienced standardised literacy testing at a 'higher' administration level, straddling both the system and site levels (Freebody & Wyatt-Smith, 2004).

The initial review of available, relevant literature allowed the researcher to narrow the focus onto the specific area of teachers' approaches to test administration. The decision to focus on the context of secondary-level, State-school testing reflects the researcher's professional area of teaching experience, as a secondary-level English/Literacy teacher. The focus on teachers' experiences, similarly, reflects this researcher's interest in the ways that teachers teach and make sense of teaching. This focus on teachers' experiences too is a result of the issues identified in the initial literature review, which indicated that teachers hold a variety of opinions on the tests, as do all interested parties, and that more benefit could come from research examining how teachers engage in testing and for what reasons.

As stated, the specific focus of this research thesis is Tasmanian secondary-level State-school standardised literacy testing. The research

includes a review of relevant literature; selection and employment of appropriate analysis approaches, data gathering and construction tools; the generation and application of the research questions grounded in the data; and the identification of findings, limitations and study recommendations. The research questions, detailed further in the Methodology chapter (chapter three), are:

Research Question 1: How do teachers perceive and interpret the value and validity of standardised literacy testing?

This question frames the research as centred on a consideration of teacher participants' understandings and opinions of specific aspects of standardised literacy testing processes and practices. This includes the ways that teacher participants express views about the value or otherwise of tests, for whom the tests are conducted, the uses of testing and test data, and the perceived accuracy and validity of test data as indicators of student literacy levels.

Research Question 2: What patterns of action and interaction characterise the classroom site of standardised literacy testing?

This question directs the research onto the classroom site of testing, to interrogate what can be seen happening at this site during standardised literacy testing sessions. 'Action' here refers to the ways in which test administration is executed, the practices and processes employed by the test administrator. 'Interaction' is taken to mean test-related interactions, behaviours and communications noted by test actors as influencing the classroom site of testing.

1.2 Study Context

This research is a small-scale case study examination. The data were gathered at one school site, a Tasmanian secondary-level State-school, through four participants (three teachers responsible for test administration and follow-up, and one principal) and two class groups (one class group each of Year Seven and Nine students, whose testing sessions were observed). The participants and observed student groups are further discussed in Chapter Three: Methodology.

The school and participants were selected in accordance with University of Tasmania Ethics guidelines and Tasmanian Department of Education (DoE) requirements. Approval was gained from the University Human Research Ethics Committee and the DoE before the site and participants were approached, and participants were given the power to veto any collected data although none chose to exercise this. These approval documents are attached as Appendix A.

This research is undertaken with the clear intention of contributing crucial information for education development, whether as pedagogical change, in practice or administration. Direct, practical and theoretical application of the research findings is an important objective. Discourse analyses can “produce recommendations for different practices and for interventions to produce change” as well as real “applications” (Taylor, 2001b, p. 325; original emphasis removed). Constructivist grounded theorists,

similarly, are expected to contribute to knowledge and understanding, in order to justify their research (Charmaz, 2005; Dey, 2007; Stern, 2007).

Further, while the research context undoubtedly influences that which is produced, and “researchers both reflect and shape the social landscape” (Witkin, 2000, p. 208), the desire for research applicability is nonetheless “a legitimate, even desirable, aspect of social inquiry” (Witkin, p. 208). The research project is aimed at developing a critical awareness of the motivations behind and effects of teachers’ perspectives and administration of, and the politics behind, standardised literacy testing. Conducting this research at a school site helps to contextualise the collected literature on standardised literacy testing, and enables an informed point of comparison for the results of other studies against a selected case study site.

1.3 Analysis and Methods

Two methodological approaches to the analysis of this textual data are employed during the research and analysis stages: constructivist grounded theory (Charmaz, 2000, 2001, 2005, 2006; Charmaz & Mitchell, 2002) and discourse analysis (Carabine, 2001; Fairclough, 2001; Taylor, 2001a, 2001b) approaches. A combination of analysis methods ensures quality and rigour of analysis, as “different perspectives provide different forms of knowledge about a phenomenon so that, together, they produce a broader understanding” (Phillips & Jørgensen, 2002, p. 4).

These analysis approaches challenge assumed beliefs and practices, and do not seek to present a wholly objective, impartial piece of research, which

would undermine the epistemological and ontological basis for the research perspectives. The analysis approaches adopted within this study instead intend to illuminate ways the texts and practices reinforce or challenge existing hegemonies (Carabine, 2001; Charmaz, 2000; Charmaz & Mitchell, 2002; Corbin & Strauss, 1990; Haig, 1995; Rowan, 2001; Taylor, 2001a, 2001b).

Constructivist grounded theory is an approach to research and analysis that deliberately draws the theory directly from the data, rather than seeking data for purposes of proving or disproving a preconceived hypothesis or theory (Charmaz & Mitchell, 2002). A grounded theory approach to data collection and analysis works from the theory generation rather than the theory verification perspective, and as such is an inductive method of developing or finding theory (Punch, 1998).

The employed analysis approaches inform, and are informed by, the researcher's acceptance of subjectivities and multiple realities, and the understanding that meaning and experience are constructed through language and discourses. Discourse analysis, as employed in this study, is concerned with the broader issues and themes across the data that represent the more overarching patterns of power, meaning-making, and interpretation, for instance. This analysis involves a questioning of the ways that the phenomena and practices reinforce or challenge hegemonies and norms, gaps, silences and counter-discourses, of the social implications of the dominant discourses in broader power and education contexts, discursive representations and sites of contestation (Carabine, 2001; Fairclough, 2001; Phillips & Hardy, 2002; Rowan, 2001; Taylor, 2001a, 2001b).

The research questions do not direct the research or analysis, which would detract from the value of the inductive nature of the constructivist grounded theory analysis. Rather, the research questions, addressed in this manner, provide a structure or orientation in the presentation of themes, patterns and dominant discourses.

1.4 Structure of Thesis

The chapters into which this thesis document is structured serve distinct purposes for the research. Chapter Two: Literature Review positions the study within the context of the relevant research, presenting and reviewing the literature and research and clarifying the knowledge-based influences upon the study. Chapter Three: Methodology explains the research questions and clarifies and details the methodological and analysis decisions made through the study.

Chapter Four: Coding Analysis of Data presents the initial results, of the first constructivist grounded theory coding stages, concluding with the identification of the themes across the data. Chapter Five: Research Questions and Themes presents the results of the final constructivist grounded theory coding stage of the analysis, through which the themes are examined and structured alongside the research questions and the discourses identified. Chapter Six: Discourse Analysis presents the findings of the discourse analysis, orienting these and therefore the data also within the educational context of broader power structures, site and system tensions (Freebody & Wyatt-Smith, 2004; Williamson & Myhill, 2008; Wyatt-Smith, 2008) and other research and

literature. The use of the terms ‘results’ and ‘findings’ in regards the coding and discursive analyses is intended to demonstrate the level of analysis and contextualisation at each of these two levels.

The final chapter, Chapter Seven: Conclusion, reiterates the research findings in context, provides suggestions for research, policy and practice from the findings, and summarises and concludes the research. Following the References is the Appendix, which presents such additional detail as the ethics approval forms, consent and information letters provided to participants, and interview and questionnaire forms, and examples of coded data.

This thesis adopts research language and terminology used by those theorists and researchers whose work has particular relevance to this topic. For example, ‘codes’, ‘categories’, ‘themes’ and ‘discourses’ as terms from the constructivist grounded theory approach serve in a variety of ways, including structuring the thesis and analysis. This research has been influenced by concepts and ideas variously described and labelled by key writers, such as Charmaz (2006) whose work on patterns and concepts relate to elements in this research, and Carabine (2001) and Fairclough (2001) who explore concepts of absences and silences, effects, tensions, and struggle. These different research terms will be explained through the text where introduced.

Through this work, Tables and Figures are provided where necessary, though their number is deliberately minimised. Instead, the focus on dialogue in this research represents teachers’ talk, and engages the reader directly with the text and the dialogue explored. The research presented in this thesis is a socio-political analysis more suited to language-based exploration of themes

and issues than to numerical or diagrammatical inserts, except where Tables and Figures present basic frequency counts, lists and summaries of coding levels.

Emphasis is shown throughout this thesis as bold, italicised or underlined text. Original emphasis in all materials such as quotes and data extracts is retained, except where clearly marked otherwise. Written responses from participants included some underlining by the original author, all of which is retained. Emphasis identified in interview and test session recordings, written responses or field notes are identified as such by the researcher. For example, emphasis in speech was noted as stress on particular words or syllables.

1.5 Summary

This research identifies and analyses some of the ways in which education policy, located within broader socio-political frameworks or paradigms at state, national and international levels, can affect those at the classroom level. This study of secondary-school teachers' testing experiences in one specific locale looks to make a contribution to informed education development, including professional practice, through strengthening understanding about the nature and consequences of the standardised literacy testing being administered and utilised in the Australian school context. This research reflects specific perspectives and views on the topic, and as a small case study can make no claims of universal and enduring truth. Nevertheless, this research has wide relevance to all those involved and interested in education, in its call for a more

nuanced consideration of the experiences and opinions of classroom teachers at the school site. This research encourages a challenging of widely accepted ‘truths’ regarding standardised literacy testing within the school site, and acts as a conduit by magnifying these participants’ voices beyond the school site and into the system level.

Chapter Two: Literature Review

Introduction

This literature review chapter is structured into two main sections; in broad terms, each addresses the literature and work related to the research questions.

This chapter presents a review of the literature and work relevant to the research, helping to contextualise the research within its social, historical and political educational climate.

The boundaries imposed by the chapter structure deceptively imply that the two research questions are clearly demarked; however, the literature and research on the two areas overlap substantially. As such, some reference to relevant classroom issues is necessary in order to examine the broad topic of standardised literacy testing, just as reference to standardised literacy testing and its attendant debates is necessary in an examination of the contemporary classroom site. The chapter sections, 2.1 Literacy and Literacy Testing and 2.2 The Classroom Site, complement each other and enable the later analysis (which, to an extent, is also structured around these research questions) to be positioned through the extant literature.

The division of these two sections serves the further purpose in identifying the often overlapping and contrasting intentions and objectives of system and (school) site players. The division demonstrates the difficulty of separating broader issues from those within the school level, and confirms research findings across the literature that present arguments and considerations across both levels in the debates.

An important note to be made about this review of the literature, also pertaining to its structure, is that this review is written in a form that eases reading, rather than presented as a list of references on each issue or topic. This is in keeping with the intent of this work to emulate dialogue and interpretation, rather than to tabulate and list. The literature review acknowledges the extensive material available about literacy testing and the classroom site, and rather than aiming to exhaustively dispute or debate that body of work, the review (and the research overall) is focused on the lived experiences of actors not always considered at the system level of education. When this focus is adopted, the literature review assumes a quite different emphasis or orientation, to one that reflects site actors' criticisms of current policy and policy-driven practice. The literature review reveals how this particular set of writings drawing on research and theory is much more critical of contextual factors that negatively affect teachers' attempts at testing for quality or instructional motivations. It is at this point that ideological and political preoccupations become clearer.

2.1 Literacy and Literacy Testing

This section has a dual focus, as its title suggests. This dual focus on literacy and literacy testing better enables a more nuanced, detailed examination of the context and place of standardised literacy testing in Australia (and internationally), including locating testing issues within the current Australian literacy situation. This section very briefly overviews the topic of literacy in Australia is presented (2.1.1 Literacy in Australia) and examines the topic of

standardised literacy testing (2.1.2 Trends in standardised literacy testing), in order to fully examine the extant literature and work on the topic of standardised literacy testing, and to contextualise the later consideration of research question 1 ('In what ways do teachers account for the value and validity of standardised literacy testing?').

2.1.1 Literacy in Australia.

This research adopts the following definition of literacy as presented in the document *Shape of the Australian Curriculum: English* (National Curriculum Board, 2009), a precursor report to the Australian national English curriculum.

Literacy conventionally refers to reading, writing, speaking, viewing, and listening effectively in a range of contexts. In the 21st century, the definition of literacy has expanded to refer to a flexible, sustainable mastery of a set of capabilities in the use and production of traditional texts and new communications technologies using spoken language, print and multimedia. (p. 6)

The above definition has particular relevance to this research project for two key reasons. Firstly, adopting the officially sanctioned definition of literacy demonstrates an interest on what the system level deems valid and implicitly communicates to Australian site level educators. The standardised literacy tests considered in this research work from this definition and therefore also represent system understandings and practical formulations of theory that ultimately affect what happens at the school site.

Secondly, the official definition highlights historical developments in the field, with evolving ideas and beliefs about literacy within a wider social

context. Indeed, as Freebody (2007) noted, the changing nature and understanding of literacy is often-noted in research, and is inextricably contextual, “both expressions of social and cultural histories and projections of preferred futures” (p. 6; also Beare, 2006; Dressman, 2007; Green, 2006).

Research has clarified that a number of understandings of literacy and approaches to literacy teaching lie in apparent opposition to each other, and that educators, occupying the tension-point at which theory and practice meet, must navigate and evaluate options to select the most fitting for themselves and according to their learners’ needs (Mills, 2005; Schleppegrell, Greer, & Taylor, 2008). Wyatt-Smith (2000) termed the debate about theory and teaching of literacy and the subject of English in Australian schools an “ideological warfare” (p. 71).

2.1.2 Trends in standardised literacy testing.

Research not only notes changing literacy definitions, practices and policy, but also changes in education itself. Apple (2001), Burch (2010) and Levin (2010) are among researchers to identify that such changes in education provision and measurement as educational standardisation, accountability and comparison have garnered mixed reviews, responses and outcomes.

The debates surrounding such education trends reflect the varied stances on the form that literacy teaching and learning should take. The contrasting interpretations of educational standardisation and accountability, noted in this and the previous chapter, include variously framed and expressed arguments that these trends are necessary reactions to recent pedagogical fads

and that they increase transparency of students' literacy progress and levels and teachers' effectiveness in literacy instruction (Cohen, 2010; Donnelly, 2004; Henderson, 2005; Lubienski, Gulosino, & Weitzel, 2009). These arguments are countered by those positing that such trends of educational standardisation and accountability not only impose additional levels of control over what happens at school and what teachers do, effectively constricting teachers' autonomy and professionalism, but that the positive impact of such practices and policies upon students' literacy learning is unproven (Beck, 2008; Berliner & Biddle, 1995; Coldron & Smith, 1999; Davies & Bansel, 2007; Doecke, Reynolds, & Roberts, 2002; Hill, 2005; Hursh, 2008; Kohn, 2000a, 2000b; Parr & Bellis, 2006; Sacks, 2000). These debates about how best to teach, gauge and improve students' literacy skills are particularly impassioned because of the widespread agreement as to the importance and value of literacy skills (S. Black, 2004; Castleton, 2010; Cohen, 2010; Dugdale & Clark, 2008).

Further recognised is the instructional use of such tests, when content, design and purpose are appropriate: "The *right* kinds of high-stakes tests can both measure *and* enhance the quality of our children's education" (Popham, 2001, p. 102, italics in original; also Phelps, 2005; Tankersley, 2007). It is not only these uses of tests and testing that are problematic and debated, however, but broader trends that have accompanied the linking of 'stakes', which create and contextualise other problematic elements of high stakes standardised literacy testing in Australia.

These trends of accountability, standardisation and comparison can be related back to a growing market-driven ideology that has not affected

Australia in isolation. Cohen (2010) posited that governments and media of other Western countries have echoed Australian concerns about the quality of education provision and outcomes, and have drawn similar conclusions about the importance of literacy evaluation, particularly for economic and employment purposes.

[The] neo-liberal model of education in Western democracies starts with the assumption that school practices are already equitable and fair and that a better education for all is just a matter of a better management and efficiency, improved standards and teacher accountability. (Kostogriz, 2008, p. 321)

School site research has indicated that educational change has had notable repercussions for teachers' roles and work, exacerbated in cases of perceived top-down mandates, time constraints and teacher resistance (Churchill, Williamson, & Grady, 1997; Muir, Beswick, & Williamson, 2010). In this way, policy ostensibly aimed at improving student literacy standards and educational provision can be seen as affecting teachers in various and unforeseen ways. According to Ladwig (2010), although such changes in schooling responsibilities and monitoring as exemplified by standardised literacy testing have only been initiated in Australia since the 1980s and 1990s, since then they have had noticeable impact and mixed reviews.

International responses to a real or perceived literacy crisis include England and New Zealand, which have had a particularly strong influence on Australian education development as 'like' countries exploring options in education. These countries serve as examples of the popularity of governments' championing of "school autonomy, diversity and choice, private

sector involvement, privatised provision and accountability mechanisms” (Gordon & Whitty, 1997, p. 453; also Fiske & Ladd, 2000). Of relevance to this research is an inherent tension as exemplified within United Kingdom education policy between the provision for teachers’ reflective practice and “bureaucratic and political imperatives to control and those imperatives win out in the actual implementation documents”, with the conclusion drawn that “in practice teachers will feel that their own reflection is considerably censored rather than legitimated” (Coldron & Smith, 1999, p. 305). As Caldwell (2010) and Lingard (2010) have argued, however, this English model (and its application to Australia) is an example of tension, rather than success, and is one that Wearmouth (2008) argued is an education system struggling to reconcile contradictory models of policy and pedagogy regarding literacy teaching and monitoring, perhaps one gradually moving away from high stakes.

Arguments about not only whether a literacy crisis is real but also about what educators and governments should do to avoid or address one are often heated. Kosar (2005) and Purpel (1999) are among numerous examples of the level of debate these topics attract. Kosar argued that standards-based reforms and testing are unpopular on both sides of US politics for different self-serving reasons, and that it is the parents who accurately identify declining literacy standards and want action. Declaring “Test scores are a lightning rod of controversy” (Kosar, p. 31), he claimed that rather than debating the use of test scores, the focus should be on the fact that not only are the generated scores accurate but that what they indicate is highly alarming. In contrast, Purpel has

likened widespread assessment initiatives to attempts by those in power to deflect attention from inherent social and economic inequities and inequalities (also Apple, 1993, 1996, 2001; Ohanian, 1999), in order to maintain such structures while also “ranking and judging children” (p. 66) through “euphemistic” (p. 64) sleight. While disagreeing on efficacy and responsibility of initiatives, researchers and proponents on both sides of these arguments do appear concerned about student literacy levels and how best to improve them (Doecke, Reynolds, & Roberts, 2002; Phelps, 2003; Kohn, 2000a, 2000b).

2.1.2.1 *Market ideology and neoliberalism in education.*

Within the Australian context, according to numerous writers (for instance, Harris, 2005; Sherington & Campbell, 2004), a pervasive neoliberal, market ideology seeks economic or market solutions for perceived social issues that were previously seen as government responsibilities. The extension of economic or market involvement to the field of education was not only the result of a shift away from ‘welfare state’ ideology; as explained by Davies and Bansel (2007): “Economic productivity is seen to come not from government investment in education, but from transforming education into a product that can be bought and sold like anything else” (p. 254). As a result, neoliberalism and market ideology translated into the education realm through policies of school funding, choice, competition and accountability. Hill (2005) summarised that, in Australia and internationally, neoliberalism in education manifests through:

deregulation and decentralisation; the importation of ‘new public managerialism’ into the management of schools and

colleges and education services; a fiscal regime of cuts in publicly schooling and further education services; commercialisation of and within schools; the charging of fees; outsourcing of services to privately owned companies; and the privatisation and ownership of schools and colleges by private corporations. ... One salient policy is employment policy: attacks on and downwards pressures on workers' rights and conditions, and on trade union rights. (p. 259)

According to Collin and Apple (2007), neoliberalism can be identified on a global scale, and has utilised ideas of globalisation in order to normalise its tenets and justify its dominance. This “neo-liberal globalization”, according to Lingard (2010, p. 141), “promotes markets over state and regulation and individual advancement over the collective good and common well-being” (Lingard, 2010, p. 141). In education, this approach is not only couched within “globalized discourses and economic concerns” (Lingard, 2009, p. 235), but also clashes with “the logics of practice and pedagogical habits of classroom teachers” (Lingard, 2009, p. 235). In spite of an “overriding and negating [of] deeply held values of professional practice” (Davies & Bansel, 2007, p. 258), and what Grant (2009) described as “no shortage of evidence for the long-term damage being wrought on students and their communities by such top-down methods” (p. 19), “neoliberalism is nevertheless widely taken up as natural and inevitable” (Davies & Bansel, p. 258).

Either because of a popular belief of the inherent value in accountability as part of neoliberalism, a recognition of the merit of testing for information purposes, or both, the idea of using tests in education as an accountability tool has bipartisan support in Australia (Bamford, 2010) as it

does overseas (DeBray-Pelot & McGuinn, 2009; Hamilton, 2003). Lingard (2010) located a reliance on populist politics in a further manifestation of neoliberalism in education, meaning a preferencing of popular demands rather than considering lessons learned from research or policy implementation. Disregard of research in policy formulation (Grant, 2009; Levin, 2010), which has been criticised in the literature, can be seen through government moves towards a choice and accountability agenda that lacks support from educators and education researchers, who point to negative repercussions of similar reforms overseas (Au, 2008; Berliner, 2001, 2002; Harlen, 2005). Rather, Lingard (2009) argued that “to date, social policy, including education policy, still appears to be framed by a neo-liberal social imaginary”, despite referring to this neoliberal imaginary as “failed” (p. 236). Accompanying educational neoliberalism and market ideology has been the blaming of teachers for student learning outcomes and, according to Kostogriz (2008), “the failure of governments to address educational marginalisation and other forms of social injustice” (p. 322). Davies and Bansel (2007) identified that “Neoliberalism strongly reinforced the undermining of the teachers’ authority that had been established with progressivism, shifting authority away from both students and teachers to state curriculum and surveillance authorities” (p. 256).

Within this surveillance that has become a key component of Australian education reforms (Davies & Bansel, 2007; Soler, 2002), one element is externally mandated, high stakes literacy testing (Doecke, Reynolds, & Roberts, 2002). Not only are Australian students’ literacy and numeracy ‘standards’ uniformly tested and measured, but they are published, compared

and used to draw various inferences (Lingard, 2009, 2010). The Australian Federal Government's 'My School' website was developed for the purposes of making comparative school performance data available for all. Proponents promote this open access as empowering for parents and the general public, while critics argue that teachers are deprofessionalised, comparisons actually narrow the curriculum and student achievement gaps are widened as a result (Lingard, 2010). Among critics of what has become high-stakes, standardised tests there is consensus that the tests inadequately highlight or consider differences in student circumstances that contribute to results, now seen publicly as educational outcomes (S. Black, 2004; Lingard, 2010), however adequately they might serve other purposes, such as highlight succeeding and failing schools (ACARA, 2010b, 2020c; Mills, 2008).

The improvement of student learning outcomes such as literacy standards has been used as justification for accountability and progress tests. This development was driven in large part by a perceived crisis in literacy, not only in Australia but internationally. Governments have been seen to initiate reforms that ostensibly address educational inequalities, although critics have identified flaws in the administration and consequences of such reforms (Freebody, 2007). International public policy comparisons of contextual issues and identified repercussions are relevant to this research. The UK and US, leaders in market ideology-driven educational reforms (Ainley, 2004), are particularly salient overseas models that provide relevant examples for Australian research and policy development. International comparisons of

education policy context and experience are illuminating, especially in terms of possible repercussions of crisis-driven government responses.

In the UK, concerns about literacy standards led to public and government criticism even of teaching methods. Soler (2002) described that a popular belief formed, “that the teaching practices needed to be changed and teachers’ behaviour more carefully monitored through the introduction of national testing. Articles in the national newspapers pointed out that without national tests there it was not possible to provide genuine comparative results” (p. 5). In addition, a new partnership between ‘public’ and ‘private’ took shape, through which calls for improved services and outcomes created avenues for business involvement in providing materials and evaluating outcomes.

[Policies] which announce ‘zero tolerance of underperformance’ and intervention in under-performing schools ... provide opportunities for replacement and/or remediation of ‘failing’ or ‘weak’ public sector institutions. The education businesses can sell school improvement – offering schools ways of accommodating themselves to the demands of state performativity and the production of new organisational identities. (Ball, 2009, pp. 84-85)

Ball (2009) also identified this trend in the US, where “state policies can create incentives and pressures for public sector providers to use private sector services” (p. 85) as part of a range of moves to address perceived crises of literacy and improve schooling. Harris (2005) argued that school-business partnerships were required; “the economic and instrumentalist imperative has

been the driving force. Performativity and accountability in terms of outcomes are privileged over everything else” (p. 75).

Direct business involvement in education has not entirely been dismissed, although neither has it been enthusiastically embraced (Burch, 2010). Skelcher, Mathur and Smith (2005) maintained that new approaches to public policy in schooling have also created spaces for action and improvement, direct accountability and a departure from neoliberalism. Harris (2005) presented the idea that education partnerships indeed work towards as well as against neoliberalism, and suggested that detrimental effects of neoliberalism upon those at the school site can be reduced. In contrast, however, are arguments (Demaine, 2005) that UK education policy influenced by neoliberal and market ideologies have not only been unsuccessful at addressing social inequality, but have absorbed some of the more damaging characteristics of market and competition.

The US has also, in recent decades, initiated a series of education reforms culminating in similar market and neoliberal measures that directly affect schools (Apple, 2001). That Australian educators and policymakers have been influenced by American reforms is evident in Australian education research that compares and contrasts the two countries, such as Freebody’s (2007) *Literacy Education in School: Research Perspectives from the Past, for the Future*. Freebody’s thoroughly researched report provides a salient comparison between the Queensland ‘Literate Futures’ policy and the often-referenced US education policy, ‘No Child Left Behind’ (NCLB). Freebody’s noting in that report of the exclusion of educational research in the US because

of a supposed lack of scientific rigour has also been noted by other researchers (Henderson, 2005; Shaker & Ruitenberg, 2007). Freebody extolled the importance of education policy being based on a realistic understanding of educational context, needs and practice goals.

Although reminding of the US's long legacy of market ideology and neoliberalism, the benefits of which they defended, Levin and Belfield (2003) contended that increasing market-influence on education through such measures as voucher programmes, charter schools and tuition tax credits were initiated ostensibly to benefit education systems and consumers – students and parents, and tax-payers.

The motivations behind these approaches were as follows: to provide greater freedom of choice of schools as a right and more alternatives for families as a response to the increasing uniformity of schools, to use market competition to make schools more effective with given resources, and to improve options for students in economically and racially segregated public schools. These solutions were also consistent with the general movement toward less reliance on government and greater reliance on markets and other forms of decentralization. (Levin & Belfield, p. 184)

These relatively recent changes presented in the US as education reforms have attracted considerable debate (Berliner & Biddle, 1995; Levin & Belfield, 2003). Again, research (Abbarno, 1998) dates trends towards greater accountability, competition and standardisation to the perceived crisis in, and of, literacy, whereby governments began to absorb and echo business alarm about perceived growing illiteracy in potential employees. As with other

market-influenced initiatives, researchers concur that a perception of increased illiteracy and connection between literacy levels and the workforce have gendered considerable debate.

The literature demonstrates that the further argument that increased market and economic involvement in education could improve poor educational outcomes has been similarly controversial. Proponents have posited an apparent failure on the part of 'welfare state' policies in addressing issues in and of education (Collin & Apple, 2007), and asserted that US market-driven education policy "combines the unassailable aim of public school accountability with a remarkably liberal emphasis on the educational interests of under-privileged and disadvantaged children attending public schools" (Swyer, 2004, p. 213). Political support in the US for neoliberal and market-driven initiatives of high stakes and accountability testing have received broad political and public support, according to Hursh (2007). Critics, however, have counter-claimed an inherent unfairness in 'choice' and 'competition' arguments, with findings that those who need more attention and empowerment can least achieve it, intensifying inequalities (Bernal, 2005; Butler & van Zanten, 2007; Demaine, 2005; Lingard, 2010; Merrett, 2006). Hursh (2007) pointed to evidence disputing claims that standardised testing and accountability and choice measures are even intended at all to empower parents and students and address current educational inequalities.

If the UK education context can and should be considered more a warning than a goal for Australian educators and policy-writers (Lingard, 2010), the conclusion could be reached from the literature that the US

education context provides similar cause for reflection. Shaker and Ruitenberg (2007) urge awareness that “US policy, emerging from contemporary neoliberal and technocratic viewpoints and funded and propagated on a large scale, has the potential to influence international thinking on education research” (p. 207). Oppositional forces against revision, however, could be considerable, as Henderson (2005) argued that within the Australian context, “the neo-conservative discursive field privileges its claims in the ‘rationality’ of tradition, while assuming the authority to dismiss alternative views” (p. 312).

Current Australian education policy directions are tending to emulate neoliberal, market-driven policies and programmes from the UK and the US, which Comber and Nixon (2009) contextualised within similar trends notable across the human services broadly and Wyatt-Smith and Klenowski (2010) noted across various institutions and corporations. Bamford (2010) argued that Prime Minister Gillard’s, and that of previous Prime Ministers Rudd and Howard, particular fascination with education policies trialled in New York show the Australian Federal government as unaware or unconcerned about the ambivalent evidence. With some frustration, Bamford wrote, “The Education Revolution is based on ideas that have already been tried, and have failed, in the United States. ... Both the major parties’ policies are set on the same goals: the standardised testing, ‘parental choice’ and school ‘accountability’ that one of its principal American proponents [Diane Ravitch] has renounced” (n.p.).

2.1.2.1.1 *Accountability.*

Although located within a managerialist paradigm, accountability measures are not necessarily intended to have altogether negative consequences.

Accountability testing can be undertaken in such a way as to directly benefit teachers and students; Popham (2003a, 2003b) advocated appropriate test design, content and use in order for best outcomes for test actors, parents, government and the public. Just as student assessments must fulfil various criteria in order to be valid and useful for all those who utilise them both at and beyond the school level (Freebody & Wyatt-Smith, 2004; Whitehead, 2008; Wyatt-Smith & Campbell, 2002), so school-wide assessments must serve often competing goals of informing teachers and the public (Wyatt-Smith & Klenowski, 2010). Various researchers have argued that it is at this point, at which school assessments become utilised for more far-reaching purposes and become “external validations of learning” (Lesch, 2007, p. 2), that accountability moves through testing become high stakes and can be seen as having either positive or deleterious effects.

Learmonth (2005) identified that the adoption of business management techniques into international education policy and practice has accompanied a focus from administration to management and attention to performance or performativity, reflecting a wider market-influenced interest in oversight, performance and accountability. What has been termed “new public management” (Andersen, 2008) and a “new managerialism” (Leithwood, 2001) has, according to those and other researchers, affected schools and educators through a changing conceptualisation of what counts as school

management, and how to ensure school improvement, though both managers and site actors are apparently subjected to scrutiny and oversight:

This 'new culture' of 'New Public Managerialism' entails a complementary and increasing governance of management bodies on the basis of mistrust targeted at both students and staff, resulting in an overkill of bureaucratic regimes of control and regulation. Formal assessment exercises require teachers to produce explicit 'learning aims and outcomes'. Students have become more subject to accountancy versions of educational values. (Hill, 2005, p. 276)

Public and government calls for educational accountability in Australia most often urge more accurate accounting of school progress or failure, with readily and publicly available reporting, upon which important decisions can be based and parents can be informed about educational outcomes and issues. In this line of argument, reflecting what Hodgetts (2010) terms "the culture of performativity pervading education systems" (p. 29), teachers have been increasingly considered responsible for student performance. The literature shows that teachers are often positioned in a tension-point between public and government expectations for performance, at the same time expected to maintain professional identities and positive relationships with students (Comber & Nixon, 2009). Mills (2008) saw this development as 'unfair':

[In] education, the accountability movement has sought to place the responsibility for instructional outcomes unfairly on teachers, seeking to apply these business management techniques and performance-based measures to complex educational contexts. (p. 212)

That large-scale, high stakes standardised testing has been seen as an accountability tool across education systems, particularly in the US and the UK, has already been noted. This subject has been accorded considerable attention by researchers, educators, media and government in those countries and in Australia. The centring of the argument on practices to produce measurable comparisons, in the form of added bureaucratic management, rather than focusing on actual underlying causes for differences, argued Harris (2005), can come at the expense of “more reflexive and innovative working at local level” (p. 83).

The need for informed decision-making regarding this contentious accountability issue has been articulated. Afflerbach (2009a, 2009b), while noting the potential reliability and validity of such testing, also maligned what he perceived as a lack of “informed research” and “scrutiny” (2009a, p. 467) as to its benefits for schooling. Indeed, Black (W. R. Black, 2004) argued that accountability measures might ostensibly encourage and promote more equitable schooling practices and outcomes, but so too might they discourage and restrict such aims in their interpretation from the system to the site level (also Apple, 2001).

The ways in which Australian mandated standardised literacy tests are used, publicised and perceived have been variously criticised in the literature. Hill’s (2005) descriptions of negative effects of neoliberalism and school ‘management’ have considerable support. In addition to mandated national tests, the current form of which is the ‘National Assessment Program – Literacy and Numeracy’ (NAPLAN), moves towards accountability in

Australian education utilise the Federal government's 'My School' website, promoted as a tool for accountability. 'My School', like the tests and test results it publicises, has not been free from debate. In this regard, Australian high-stakes testing continues to be subject to the same active debate among researchers and governments as in the UK and US, fuelled by media reporting and lack of clear scholarly consensus.

The Australian media plays an important role in promoting the goal of educational accountability as a means of widening parental school choice and improving educational outcomes. For example, the following extract from a *Sydney Morning Herald* article (Buckingham, 2003) on test-oriented educational accountability shows popular media reporting making wide claims for benefits and consequences:

Greater public scrutiny would increase competition, improve education and cut costs. ... Accountability is not a benign concept, and simply making information available is not sufficient. There must also be consequences. In an open education market, it is relatively simple: accountability is achieved through competition. A school that fails to provide what is reasonably expected would lose students and therefore funding. (pp. 1-2)

The deliberate linking of school improvement with accountability, competition, scrutiny and funds – as well as strong language highlighting stakes and assessment itself ("fails to provide") – reinforces the pervasive market ideology and performance culture within which literacy testing is located in Australia.

Calls for educational accountability, with the inherent implication that previous education provision lacked appropriate accountability measures, have been persuasively argued with reference to parents' and taxpayers' rights to be fully cognisant of what is being learned at school; this information is presented, and considered, as a justifiable return for public funding and trust. Currently, NAPLAN test results are publicised through government press releases, posted on the 'My School' website and distributed across public media outlets. Analyses of test results have tended to highlight score disparities, especially indications of supposed failures and identification of individuals and groups who could be held responsible.

Through his paper *Accountability and the Public Purposes of Education*, Reid (2010) reviewed the ways in which current 'My School'-based accountability measures fail to achieve the stated government intention of "transparent accountability" (p. 3), described as "information for parents choice; information for quality improvement; and information to inform policy action" (p. 3). Reid echoed other researchers' arguments that the concept of parental choice in children's schools is inherently misleading and misinformed, as many parents cannot in fact 'choose' at all, due to location, cost and school zoning, and that tests and their comparisons do not and cannot explain differential educational outcomes. Reid concluded that 'My School' "should be shut down now while further development work is done. I don't think it can be patched up as we go, with the fitful addition of disparate pieces of information as lobby groups argue their case" (Reid, p. 7).

Other classroom- and teacher-focused research has produced a range of findings that appear to support Reid's (2010) assertions. High-stakes accountability tests have been found to have "unanticipated, and often negative, consequences for teachers' relationships with students, pedagogy and sense of professional well-being" (Valli & Buese, 2007, p. 519), and that often principals and teachers find utilising or interpreting test results difficult, enabling public "overreactions to what may well be hyperrationalized analyses of small differences in moderately reliable measures over time" (Stringfield & Yakimowski-Srebniak, 2005, p. 68). Furthermore, Booher-Jennings (2005) found that teachers, under pressure of being held accountable for student test results, can resort to behaviours that undermine the accuracy and usefulness of test data, with teachers' classroom approaches to teaching and their very professionalism suffering (Webb, 2005).

2.1.2.1.2 Deprofessionalisation.

Teacher deprofessionalisation as one potential consequence of high-stakes literacy testing has been raised in the literature. As noted, accountability measures have been seen as leading to narrowed curriculum, constricted pedagogy, performance-based pay and teacher blame for poor student results. These concerns about teacher deprofessionalisation are not limited or specific to the Australian context, and Hill (2005) posited that these are justified concerns in neoliberal education systems.

Doecke, Reynolds and Roberts (2002) noted that teacher deprofessionalisation as resultant of neoliberal and accountability measures in

education policy and practice is disputed by government claims that the educational objective is in fact improvement of the status of proven successful teachers. Yet government rhetoric and political intent in education policy development has too been questioned in Australia, as elsewhere. In his examination of such education reforms, Beck (2008) found that the English government adopted, subverted and manipulated understandings of teacher professionalisation, in “an attempt to silence debate about competing conceptions of what it might [mean] to be a professional or to act professionally. The overall process is thus arguably one of de-professionalisation in the guise of re-professionalisation” (p. 119). Brown’s (2008) research into standards-based accountability reforms in Wisconsin indicated that “Stakeholders linked accountability with transparency” (p. 277), even though that state’s Attorney General publicly expressed her criticisms of NCLB and its implementation as in some regards unconstitutional and illegal (Dwyer, 2004).

Research in the Michigan context also indicates complexity in school-site administration of accountability measures. Gawlik (2009), for instance, demonstrated that although teachers and principals in Michigan charter schools were at times frustrated and overwhelmed by accountability requirements, they did not altogether resist accountability in schools. Other research findings report similarly ambivalent teacher reactions to changes in school accountability measures. According to Webb (2005), teachers he studied including in the US at times accepted and “appropriated” (p. 204) apparently deprofessionalising accountability requirements, though he warned about the

divisive deprofessionalising potential of such neoliberal accountability measures, especially when teachers' expertise is overlooked and ignored, "the very people held accountable for the issues" (p. 205).

The literature indicates that neoliberal and market-influenced teacher accountability and management measures have notable consequences for what teachers do. Comber and Nixon (2009) particularly noted the pressure placed on teachers to shift attention from a pedagogical and teacher-student focus to meeting external requirements. Again echoing the performativity element, Webb (2006) located external expectations – "expected test scores, expected norms for student behaviour, expected exemplifications of curricular prowess" (p. 211) – and performance within demands for transparency and accountability that are sourced beyond the classroom and educational environment. Such performance, he argued, takes the form of "inspections, observations and public performance of test scores" (Webb, p. 201).

Undermining of teacher authority and professionalism that can accompany increased accountability demands and external reporting is evident according to some writers, through "shifting authority ... to state curriculum and surveillance authorities" (Davies & Bansel, 2007, p. 256). This shifting of authority from those who teach to those 'outside and above' the teachers – from those who know about their students' needs to those who might not know (Webb, 2005) – can silence or suppress teachers' classroom experience and views, and curtail their practice (Prashad, 2006).

The concept of performance-based or merit pay for teachers has been the subject of a similar level of debate within the research on accountability

and related market-driven controls. As argued, performance-based teacher pay was ostensibly proposed as an incentive for teachers to be more productive, as a reward for improved student learning outcomes and as encouragement for younger teachers to stay in the profession. In Australia as elsewhere, as Cole (2010) and Naidoo (2010) reported, many teachers and teacher groups dispute these motives and challenge the concept of performance-linked remuneration, questioning the basis for decisions about teacher performance, and pointing to alternative approaches to improve teacher retention and recruitment. Coleman (2008) and Miner (2009) asserted that proposed performance-based pay insults teachers' professionalism with the implied assumption that financial enticements can somehow increase teacher effectiveness, although a recent OECD report identified teacher pay as highly affective upon teaching and learning outcomes (OECD, 2010).

It is also argued that other externally designed and measured test data, such as teachers' assessments, are already available and could be utilised (Reid, 2010). Government insistence on their preferred choice of data (Booher-Jennings, 2005) could reflect Freebody and Wyatt-Smith's (2004) recognition for system validity in educational data, but could also communicate a public or government mistrust of teachers' mediated assessments.

High stakes literacy testing, then, although ostensibly serving the goal of providing more transparency in school outcomes and clearer data about student literacy standards, serves other goals also. Luna and Turner (2001) noted that in addition to students and teachers, increased calls for site level accountability and performativity ultimately has high stakes for the entire

education system “that faces a political movement that ignores what educators and researchers know about literacy, learning, and equity in education” (p. 87).

2.1.2.1.3 School comparison and competition.

Market ideology in the education arena, focused on direct school comparison and competition, has previously been raised. In Australia and elsewhere, this is achieved by publishing school ratings and comparative status through such mechanisms as online channels and public media outlets. Proponents of school comparison and competition present objectives of seeking to improve school and student progress by encouraging transparency (see Gorard & Fitz, 2006); these arguments echo business claims of market choice enabling poor- and high-performing schools or teachers to be identified, and appropriate responses to be taken. While Lubienski, Gulosino and Weitzel (2009) summarised the general market ideology claim with their prediction that “competitive incentives will change the organizational behavior of schools (and districts, dioceses, etc.) in ways that will lead to more equitable access for students across varied and often segregated urban landscapes” (p. 601), Bracey (2008) insisted that the reality of school comparison and competition for schools – for teachers and students, disadvantaged or otherwise – is less defensible and “much more complex than those who would hold schools alone accountable ... would have us believe” (p. 621).

Merrett (2006), supporting the argument that market forces have resulted in harm for those schools and students least able to compete, asserted that popular ideas of blaming schools and students for lower literacy standards

do not hold up against overwhelming evidence that “a child’s background is still the major determining factor in their educational achievement” (p. 94). Merrett further posited that student relocation to schools with higher average socioeconomic status (SES) of students “is unlikely ... to break this link” (p. 95) and that, instead, market ideology in education has been, and will continue to be, manipulated by parents of higher SES students.

Market forces operate through competition and create winners and losers, designating schools as effective or failing as a natural consequence ... [Policy makers] should stop blaming schools and school leaders for urban failure, admit the failures as well as the advantages of market forces in education, and work to find ways of addressing social values and attitudes at an earlier age. (Merrett, 2006, p. 96)

Bernal (2005) and Merrett (2006), too, argued that trust in the overall justness of market initiatives in education is misplaced as they widen, rather than narrow, class divisions, particularly as competition enables schools to be labelled as winners and losers and poorly-performing schools lose high-achieving, mobile students to schools designated as high-achieving (also Fiske & Ladd, 2000). Bagley (2006) and Lynch and Moran (2006), too, noted that parents and schools actively perpetuate consumerism and competition for (personal and competitive) advantage.

Despite reservations by researchers and disagreements at times about ways to respond to the ‘crisis’ and methods to achieve optimum results for all students, the concept of school comparison and competition is promoted as an answer to widespread concern about failing schools. The reasoning is that, since the fault lies at the school level, those at the school level cannot be

trusted to identify an appropriate solution, particularly sufficiently “scientifically based, technical solutions” (Cohen, 2010, p. 112).

Thus it is that ‘scientifically based’ solutions come to shape such important education policy as NCLB, despite “ethnographic and other qualitative research methods that have become widely valued by teachers and administrators” (Shaker & Ruitenberg, 2007, p. 210). Even if the policy intent of testing does not deliberately involve decision-making about school rankings, nevertheless when the public welcomes such outcomes these unintended consequences can gather momentum. Levin and Belfield (2003) posited,

Market approaches increase choice considerably; [additionally,] competition and choice are associated with small improvements in academic achievement, but nothing approximating the revolutionary changes argued by advocates. (p. 212)

Mills (2008) and others have pointed out that these tests used to compare schools and encourage overt school competition are not only imperfectly designed for the purpose, but that other available evaluation methods could be more appropriate for the same outcome. Amrein-Beardsley (2008) and Luna and Turner (2001) report concerns that single tests are used to judge students or systems, or to make wide-reaching decisions, and Ohanian (1999) posited that teachers and parents already know student literacy levels – that a focus on standards is not intended to narrow schools’ achievement gaps but to widen and justify them. Apple (1993), too, argued that rather than undertaken from any laudable aims, standards-based initiatives (such as high stakes, standardised literacy testing) can be seen in a broader pattern of control of, control over, and access to knowledge. Even if superior tools were utilised to

more accurately facilitate school comparison and competition, the literature nonetheless persuasively demonstrates the need for more convincing justification.

2.1.2.1.4 Economy and employment.

Collin and Apple (2007) described “the ongoing and conflict-ridden shift ... from welfare state industrial economies to neoliberal state informational economies” (p. 434), in which focus is consistently turned to ways that “workers add or learn to add value in economies driven in powerful ways by the generation, assessment and application of new knowledge” (p. 434). This is the contemporary context that shapes the education environment at all levels and locations. The influence of market ideology in the education arena, as previously noted, has involved arguments that high literacy standards are beneficial for economic and employment reasons. This political and economic framework invites images of students as future contributors to the workforce. Education comes to be seen as a preparatory mechanism for achieving future workers who need to be aware and capable of assuming their essential roles as producers and consumers, willing actors in a neo-capitalist economic system.

A solid body of research notes the connection, and the nature of the relationship, between education and literacy (among other skills) on the one hand, and the economy and employment levels on the other. The benefits of education and literacy identified through research include increased job security, income, health and wellbeing, and political involvement (for instance, Castleton, 2010; Dugdale & Clark, 2008). Further, in the context of

developing nations, education is seen as an investment for “national development and poverty reduction” (Johnson, 2010, p. 186). Minority, disadvantaged groups are particularly seen to need education for citizenship participation in wider society, including access to opportunities for class mobility which can multiply benefits in turn, as a trickle up notion of social development.

Black (S. Black, 2004) problematised dominant Western perceptions of the link between lower literacy levels and unemployment, and public belief in economic arguments regarding economic and employment attitudes to literacy, challenging basic premises promoted by government and industry concerning the relationship between literacy and economy/employment.

Lack of literacy and numeracy skills do not cause unemployment or limited employment opportunities. Rather, it is these economic conditions that cause literacy and numeracy problems. They give rise to the need for governments and others, acting in the interests of capital, to introduce competition policies and shift responsibility for productivity and efficiency to workers, to produce and inflate literacy and numeracy problems, even though, as an increasing number of ethnographers of workplace indicate, literacy and numeracy are not the problem, or at least not the main problem. (S. Black, 2004, p. 15)

Black's (S. Black, 2004) view is supported by Collin and Apple (2007), who indicated that current high-stakes literacy tests in the US deflect public and government focus from what educationists believe is most needed, and instead reinforce current economic priorities. Australian governments and industry have readily embraced neoliberal solutions to perceived social problems; in

this context, as already noted, government and industry have been at the forefront of calls for schools (and for those who work within them) to be subject to the same competitive, market pressures as private enterprise. Murphy (2009), for instance, undertook research suggesting, “the literature on turning around failing organizations in sectors outside of education provides potential blueprints for recovery activity in failing schools” (p. 796).

Despite its uptake, this ideology has been received with ambivalence and contention. Within the literature is both significant support of and opposition to government-run schools being treated as private enterprises, subject to market whims, competition and labels of winner and loser. Such treatment does arguably fail to recognise the role that schools play in communities and the importance of education for reasons other than economic. Business- and market-influenced claims that educating children has economic benefits has led to much discussion and questions about investment in education, including fears about the impacts should education lose its status as economically valuable. Johnson (2010), for example, argued that the current government and industry support for education could be withdrawn “if data suggest that it is more economically efficient that particular groups of children” not be educated (p. 210; also Tomasevski, 2006). A further and potentially more cynical argument made by Albert (2005) is that rather than failing, schools are in fact succeeding in their actual purpose, which is to develop and train workers and perpetuate a market-driven economy.

2.1.2.2 Test content, use and reporting.

The issues of decision-making regarding, and content of, Australian standardised literacy tests are subject to widespread debate. Varied opinions are expressed by teachers' groups and professional associations (such as the Australian Education Union [AEU] and independent schools' groups), government authorities and academic experts. Media headlines have reflected a degree of discord concerning the tests in Australia, for example *The Age* newspaper's 'Literacy tests full of flaws' (Craig, 2010) and 'NAPLAN tests help champion equality in the classroom' (Jensen, 2010b); *The Australian's* 'Standardised tests fail students, say teachers' (Ferrari, 2007); the *Herald Sun's* 'Union drops planned ban on literacy and numeracy tests' (Hudson & Masanauskas, 2010); the *Sydney Morning Herald's* 'Better supervision needed to stop cheats, say schools' (Patty, 2010); and *Inside Story's* 'NAPLAN and the states: an intriguing result' (Suggett, 2010).

Among concerns, media reports have particularly focused on teacher groups' and researchers' criticism of the questions contained in the NAPLAN literacy tests despite, or because of, the identification of blatant errors (Craig, 2010). ACARA, responsible for the creation of the NAPLAN tests, has refuted criticisms, in turn claiming that test content is designed in consultation with teachers, and is continually revised for improvements (Craig, 2010; Freeman, 2009).

In this way, the involvement of teachers in test design and content has represented a further issue. Given that it is teachers who administer and most immediately and directly make use of the tests in the classroom with students,

their opinions of the tests carry influence. Teachers have the capacity to effect the success of standardised testing at the (school) site level, as shown by concerned Australian and UK teachers' threats to boycott such testing (Hudson & Masanauskas, 2010; Jensen, 2010b; 'Leading article: We can't ignore the teachers' boycott of tests', 2010). As noted by Craig (2010), a common criticism levelled at test content is its irrelevance to curricula, and teachers have argued that tests must be directly relevant to what happens at the (school) site in order for them to appreciate test processes as directly relevant for their literacy teaching and interventions.

These concerns also play to the problematic issue of test validity, and testing proponents have emphasised that "[b]enefits for school and students from NAPLAN depend on literacy and numeracy tests that are thorough, accurate and objective" (Freeman, 2009, p. 13). According to Lake Corporate Consulting (2006), the validity and reliability of tests is important to any subsequent use made of the results. The validity, accuracy and objectivity of test content is widely recognised as influential on usage and follow up activities involved with high-stakes literacy testing; therefore, claims of poorly-worded or irrelevant test content are accepted as serious challenges. Such issues in test content can be intensified as the 'stakes' are raised when government reporting of school results become transformed into headline news by the public media.

Reporting of test results has opened a particularly critical avenue for debate about the tests. Jensen (2010b) argued that the AEU's longstanding opposition to high stakes testing, now directed at NAPLAN and its reporting

through 'My School', confuses the issues. Jensen posited that the "best education systems use [national] student assessments to compare the performance of students and schools and allocate resources accordingly. This helps reduce inequality" (Jensen, n.p.), and that problems perceived with online 'My School' reporting do not automatically mean the problems lie with the tests themselves.

Jensen's (2010b) viewpoint was that outright boycotting and opposition of both NAPLAN and 'My School' would paint the AEU "as anti-reform" (n.p.). Commentary about this stance seemed to be mindful of this potential, and the AEU Federal President, Angelo Gavrielatos, had stated that teachers' support of a boycott of NAPLAN and 'My School' "has been totally misrepresented as an opposition to accountability and transparency, which is regrettable" (Hudson & Masanauskas, 2010, n.p.). This statement reflects the contextual dynamics and contestation involved with reporting of results and media involvement. Teachers obviously hold a difficult position here in what appears to be political struggle over education directions, in particular accountability through means such as the NAPLAN tests.

The potential for cheating within test sessions is a major factor that influences the validity of test results. Public opinion of teachers and the tests is affected, as are teachers' attitudes to utilising test data under such conditions; high stakes attached to test results can play a part in creating further pressure. The media has been attentive to "reports of principals and teachers helping students with answers in the literacy and numeracy tests to improve their schools' rankings" (Patty, 2010, n.p.). The Association of Independent

Schools of New South Wales reacted strongly to these reports by taking particular issue with test “cheating and competition” (Patty, n.p.). Daniels (2005) pointed out that independent and private (non-government) schools’ reputations depend on their perceived academic success, as indicated by test results. Public (government) school teachers’ test cheating therefore calls into questions those tests which are the premise for independent school popularity and narrows the gap (however minimally) between public and private school test results. In this context, questions about the appropriateness of the tests used for making comparisons (Craig, 2010) become particularly sensitive. One suggested response to calls of test cheating has been to employ “greater security, including independent supervision” (Patty, n.p.), which could be seen as a move to strengthen independent, that is, non-teacher, control of the process through designing, mandating, analysis and reporting of the tests.

Publication of test results is not only problematic because of the ‘My School’ website profiles; for some teachers, researchers and commentators, their concern is with the public media involvement. “[M]edia framing has a powerful impact on public opinion ... [and] can affect educational policy-making, preferred modes of teaching and learning, school content and parenting values” (Ognyanova, 2010, p. 51). Nevertheless, utilisation of public media offers potential for communication of political purposes, messages and gains:

Such reforms are not only attractive through their use of the symbols of standards, accountability, and improved student performance; they also provide policy makers with a means to garner attention as concerned politicians and to mediate a public

institution that attracts attention of the business community.
(Brown, 2008, p. 285)

Snyder (2008, 2009) claimed that media reports of crises and reforms, and school successes and failures, not only ensure sales but enable regular exposure and consumption by the public of stories about students suffering under poor teacher quality, literacy teaching fads and falling literacy standards (the latter often directly attributed to the first two claims). Examples of media reporting include explicit naming and judging of individual teachers and particular schools (Hudson, 2010; Simpson, 2010; Patty, 2010). Despite potentially destructive reporting in relation to tests, however, there exists significant potential for positive media influence, as suggested by Brown (2008). Governments can utilise media communications to inform and shape public opinion on literacy issues, and the public media can similarly exercise its capacity to inform and shape not only public opinion but by extension progressive public policy.

Levin (2004), examining the mutual need and mistrust between media and government, noted trends of simplification and blame within media and government communications, and a tendency to focus on short- rather than long-term issues. As a result of significant shifts in accessibility of government information, further facilitated by widespread Internet accessibility, a general public once dependant on media interpretation and mediation of primary data is now able to access such data directly. Direct access is an important issue given Levin's description of prevalent public frustration and suspicion about the trustworthiness of political and media spin,

particularly salient in the midst of current Australian debate about misrepresentations and the state of literacy among students.

Reporting of test results demonstrates deliberate publication bias, according to a number of writers, for example Torgerson (2006) who wrote that “the tendency for a greater proportion of statistically significant positive results of experiments to be published and, conversely, a greater proportion of statistically significant negative or null results not to be published” (p. 89) represented “an important threat to evidence informed research and policy-making” (p. 99). Selective emphasis on preferred test results can be seen as publication bias by governments as well as (public) media outlets.

Media representations of literacy trends and scores, according to Warmington and Murphy (2004), communicate an implicit message: sensationalised education reporting reminds teachers that they have no ownership of education debates. Those authors presented the theme that teachers are unduly and pessimistically influenced by media portrayals, yet teachers are already subjected to often-conflicting expectations, even without media involvement. Benjamin (2003), for instance, in her analysis of the effect of media portrayal of school success and failure, noted an inherent tension between pushes for schools to be inclusive of students’ differing needs and what she termed the “‘standards agenda’: the top-down drive to improve students’ improvements” (p. 105). Gale and Densmore (2002), too, argued that commonplace labels defining success and failure are used by teachers, students and parents, and that these labels shape the ways that these actors teach and consider students. From such examples can be inferred that educators do

construct part of the general perception of what is meant by student and school success and failure. Arguably, however, this perception is also framed by representations communicated by, for instance, governments, industry groups, and media outlets.

Educators' perspectives at times concur with media portrayals of certain issues, and on other subjects there are major differences. On the subject of literacy, a gulf can be relatively easily identified between media on the one hand and teachers and education researchers on the other (Ognyanova, 2010). Cohen (2010) pronounced a particularly apparent gulf between media and some academic representations of a literacy crisis and teacher performance, supporting claims of teachers' lack of active ownership of education debates and advocating that educators must engage with mainstream media in order to more openly and effectively engage with these education debates.

However well intentioned, educational reporting of test scores even when apparently straightforward can prove troublesome, as in cases of test bias. Braun, Zhang and Vezzu's (2010) examination of bias through the US National Assessment of Educational Progress suggested that issues in score comparisons can be created unintentionally, for example differences in instructions given to teachers for administering tests and problems with rules of testing such as who to test, that are communicated to schools. Using the example of English score reporting, Hilton (2006) found that some cases of test reporting bias are deliberate even on a national level. In contrast to that English example of national-level manipulation of testing, Ertl (2006) noted that taking part in international PISA tests led to Germany adopting rafts of

external assessment measures and reconceptualising educational outcomes and competence, which he argued had positive effects.

One suggested approach to improve utilisation and follow up of standardised literacy test results is that of contextualised, ‘value-added’ measures of student learning outcomes that might be highlighted in the tests. Arguments in support of value-added measures focus on the potential gains from such test results that are un- or under-realised in current comparisons and statistical analyses of test data. Instead, value-added measures could incorporate consideration of students’ progress over time, which could reflect teaching influences, for instance.

It is more defensible, for example, to examine a teacher’s effectiveness on the basis of how much the teacher’s students learned from the time they entered the classroom to the time they left than by simply relying on a traditional “snapshot” measure – a measure capturing the level at which students exited the classroom independent of their level when entering. ... the best way to measure school performance is to measure the gains posted by students longitudinally – to measure the value that the district, school, or teachers added to students’ learning over time. (Amrein-Beardsley, 2008, p. 65)

Kelly and Monczunski (2007), whose research supported such a measure enabling and validating yearly progress, argued that value-added measures could ensure more thorough use of accountability and monitoring testing across students and schools by providing more accurate explanations for student scores and identification of specific influences.

Prior to the introduction of NAPLAN, many of the standardised literacy tests used in Australia were insufficient in content, design and utility for value adding, and had relied on “statistical adjustment” (Lake Corporate Consulting, 2006, p. 2) to scale each year level’s scores. While not serving as a correction for any dissimilarity written into tests, value-added analyses are promoted as a way to increase the usefulness of test data (Jensen, 2010a, 2010b).

Despite gaining popularity, value-added measures might be improvements on other approaches but have not been unproblematic in practice: “perhaps their sophistication might limit their policy use in relation to schooling policy alone” (Lingard, 2010, p. 138). Kelly and Downey (2010), too, identify that value-added measures can be useful for a number of purposes and appear to be better than alternatives, yet their actual usefulness is somewhat hindered due to uncertainty about purpose. Merrett (2006), though, reached unequivocally damning findings about value-added measures in UK practice. Gorard (2006), too, provided further cause for concern with comments that the “apparent precision and technical sophistication” (p. 241) of value-added measurements can be misleading, the consequences for which, he explained, “are legion” (p. 242). Nonetheless, Jensen (2010a, 2010b) insisted that improving uses and analyses of test data has the potential to “counter misleading and simplistic league tables that concentrate on students’ raw test scores” (2010b, n.p.).

2.2 The Classroom Site

This section extends from the works examined in 2.1 Literacy in Australia to critically review what the literature identifies as factors within the classroom site for teachers and students relevant to this research, and current knowledge about the ways that classroom and school actors are affected by and respond to educational issues. This section's examination of the literature and research more specifically related to the actual classroom site contextualises the later consideration of research question 2, 'What patterns of action and interaction characterise the classroom site of standardised literacy testing?'

2.2.1 Teachers and students in the classroom.

Hemsley-Brown (2004) identified problematics of "the accessibility and relevance of research, trust and credibility; the gap between researchers and users, and organisation factors" (p. 534) as issues affecting broader fields of management. This concern with research accessibility and relevance is seen as a key issue within the field of education. Ercikan and Roth (2006) went as far as to argue that the polarisation of educational research into such binaries as qualitative and quantitative can not only deter and discourage educators' access of the produced knowledge, but by extension can also undermine its usefulness and relevance for those who would make best use of it. They argued that the very complexity of teaching is not reflected in such restrictive approaches. Whitty (2006) expressed concern that researchers who direct their work towards the specific aim of policy utility are not necessarily producing the best-directed research because not all policy can be research-based. Whitty's view that research aimed at practitioners has merit is relevant to this research.

Educational research is framed by implicit and explicit decisions regarding knowledge, such as whose or which knowledges are most valuable (Kelly, Luke & Green, 2008; Kincheloe, 2003), an issue especially evident in debates over curriculum and standardised testing (Apple, 1993, 1996; Au, 2007; Ohanian, 1999). Similarly framed are those studies that problematise what and whose evidence is valid and important in education research (Clegg, 2005; Gorard & Fitz, 2006; Shaker & Ruitenberg, 2007; Slavin, 2004, 2008), and indeed even what is ethical in education research (Hostetler, 2005).

This section reflects the view that classroom-based research is necessary to contribute deeper understanding about classroom learning and working environments, in which teachers and students utilise roles and behaviours in order to interact, identify and engage. A substantial body of literature regarding education, teaching and classrooms extols the importance of evidence-based policy and practice that can utilise the experiences of teachers and students in order to improve schooling (Ball & Forzani, 2007; Weimer, 2006). Further studies identify that, as a gulf exists between education researchers' and classroom teachers' understandings of teaching (Bulterman-Bos, 2008; Wiliam, 2008), research should be a collaboration between these two groups with a valuing of direct teachers involvement in the generation of relevant and classroom-situated knowledge (Berliner, 2001, 2002; Kincheloe, 2003; Lagemann, 2000; O'Mara & Gutierrez, 2010; Wardekker, 2000). Research has also noted the growing calls for student involvement in education research particularly into assessment trends and practices that affect students' perceptions and attitudes regarding schooling

(Blum, 2000; Brookhart & Bronowicz, 2003; Duffield, Allan, Turner, & Morris, 2000; Hargreaves, 2004; Whitty & Wisby, 2007; Zion, 2009). The classroom site is a necessary consideration regarding educational inequity, particularly given a focus on assessments that provide snapshot summaries of results rather than detailed explanations of causes of such inequity (Alloway & Gilbert, 1998).

2.2.1.1 *Classroom behaviour as performance.*

This research draws on theorisation of classroom behaviour incorporating Goffman's (1990) concept of the everyday social actor's presentation or personification of context-specific roles. In this theorisation, teaching can be considered a form of performance, which Sawyer (2004) similarly termed "creative" or "improvisational" performance (p. 12), through which classroom actors enact various roles and present themselves in chosen or expected ways. Within this simplified understanding of classroom behaviour as performance, all of the classroom actors are both performers and audience, and engage in expected behaviours to fulfil expectations of their accepted or allocated classroom role, such as 'teacher' or 'student'.

Atkinson (2004) claimed that fields of educational study have become normalised, and that as a result, creative approaches of demystification are required. This is why the idea of presentation or performance (Goffman, 1990; Sawyer, 2004) has resonance in contemporary classroom-based research and why this perspective informs discussion of the classroom site in this section. Considering the roles played by classroom actors (in this case, teachers and

students, as well as principals where relevant) throughout the ‘performance’ can enable a more sympathetic and nuanced questioning of what happens in the classroom and school sites. Here, more recent research (Morine-Dersheimer, 2006) supports claims made in older classroom research (Delamont, 1976; Furlong, 1976; Walker & Adelman, 1976).

2.2.1.2 *Teacher-student interaction.*

The classroom site, expectations and interactions of classroom actors (here referring to teachers and students) demonstrate the changing nature of the Australian educational, learning context. Classroom teachers “work with more diverse communities in times characterised by volatility, uncertainty and moral ambiguity. Societal, political, economic and cultural shifts have transformed the contexts in which teachers work and have redefined the ways in which teachers interact with students” (Beutel, 2010, p. 77).

A wide range of research directed at improving education outcomes has been conducted in the field of classroom interactions and changes over time have been noted. There has been a strong focus on teacher-student interaction and topics within this field have varied considerably. For example, research has found collaborative classroom talk and interaction enable teachers to improve their teaching, the students’ classroom environment and learning outcomes (Parker, 2006; Sawyer, 2004; Thwaite & Rivalland, 2009). Classroom research has further indicated that encouragement and guiding, rather than discipline perceived by student as unfair and punitive (Emmer & Gerwels, 2006; Kulinna, 2007/2008; R. Lewis, 2006), have the most

constructive outcomes in terms of students' subject-specific motivation, engagement and behavioural self-correction (Beutel, 2010; Den Brok, Levy, Brekelmans, & Wubbels, 2005; Hoy & Weinstein, 2006; Pianta, 2006; Walker & Adelman, 1976).

Such a classroom management approach, too, ultimately affects the overall classroom climate (Rishel, 2005; Rishel & Miller, 2005; Van Petegem, Creemers, Rossel, & Aelterman, 2005), and can contribute to a positive, student-centred learning environment in which "teachers are more likely to meet students' individual needs and abilities" (Walker & Hoover-Dempsey, 2006, p. 665), although researchers such as Garrett (2008) have questioned how realistic and academically beneficial such student-centred approaches in fact are. Nonetheless, the impact of teachers upon students is recognised, leading Beutel (2010) in particular to argue that teachers have more of an effect on "student achievement than other factors, such as school influences" (p. 78).

An examination of classroom interactions must be conducted with an appreciation that all classroom environments, as specific cultural and social organisations, are different, and that within such environments students and teachers may reach varying interpretations or responses to, for instance, situations or interplays, and develop their own context-specific behavioural and identity norms (Eisenhart & Borko, 1993; Furlong, 1976; Walker & Adelman, 1976). Walker and Adelman's findings of the significance in classrooms of teachers' contextual appreciation and use of humour, which call on a shared history of meaning-making, experience and social identities and provide a means by which they and their students can navigate a complex, fluid social

situation, are reflected and upheld in more recent research (Berk, 2002; Torok, McMorris, & Lin, 2004; Ullucci, 2009).

Classroom-based research shows that the classroom site in which teachers and students engage is the site at which numerous aspects intersect and interplay, such as temporal context, actors' status and identity, and the physical, institutional and educational setting, and that at this site classroom actors' roles are often unequal in power relations, highly context-specific and prone to transition (Delamont, 1976; White-Smith & White, 2009; Wilson, Croxson, & Atkinson, 2006). In such a classroom context, Delamont identified the teacher as holding power and autonomy, both within and over the learning site, notably over decisions regarding access and definitions of knowledge; potential infringements to this autonomy serve as teachers' source of vulnerability.

2.2.1.3 *The classroom environment.*

Management of the classroom learning environment necessitates teacher attention to “establish[ing] and sustain[ing] an orderly environment so students can engage in meaningful academic learning [and] to enhanc[ing] students' social and moral growth” (Evertson & Weinstein, 2006, p. 4). Necessary in achieving these dual goals, according to numerous studies, is positive, constructive teacher-student relationships (for instance, Cotton, 1995; R. Lewis, 2006; Park, 2008; Sawyer, 2004; Walker & Hoover-Dempsey, 2006; Wentzel, 2006).

In Delamont's (1976) seminal research into the classroom environment, teachers were found to hold classroom autonomy and power over students, but more recent research by Park (2008) indicated that dichotomies, "distinctive roles of the teacher and the students as the regulator and the regulated, respectively" (p. 12), are not always accurate representations of classroom roles. Park found that more often than not, classroom actors can be seen to have "cooperated, negotiated, and worked in tandem with each other for the pedagogic goal" (p. 12), rather than performed according to accepted simplistic roles relating to being empowered or disempowered.

Park's (2008) findings were echoed by Hoy and Weinstein (2006) and Pianta (2006), who identified that teachers' and students' opinions of effective classroom management often differ. For instance, while students demonstrate increased motivation in a positive and non-punitive learning environment and management approach, teachers might feel pressured to prioritise academic results and task completion (Hoy & Weinstein; Pianta). For less-motivated students in the middle years of their education, Pianta's research indicated that teacher-student relationships assume greater importance, and "can actually exacerbate risk if they are either not positive or do not match with the developmental needs of the child" (p. 699).

Reeve (2006) suggested that another key factor to be considered regarding students' needs in the classroom learning environment is teacher acknowledgement of student experiences and opinions on tasks and expectations. A number of studies (Emmer & Gerwels, 2006; Nucci, 2006; Reimer, 2005; Wentzel, 2006) concluded that in consideration of student

needs, teachers need to demonstrate appropriate and expected behaviour. This is particularly because students do not necessarily come to school already knowledgeable about behavioural and academic expectations. Greene (2008) addressed the issue of expectations on teachers to develop knowledge about influences upon students at schools and ways to deal with them. Greene argued that students whose behavioural or academic performance falls short of expected standards are probably trying, but might have specialised needs. Addressing these students' needs requires an appreciation of possible reasons for their behaviour or academic outcomes. Large-scale test measures are not necessary to develop this awareness.

Considerable research has considered underlying causes of student behaviour that merit teachers' resorting to management and discipline strategies. Teachers (especially in secondary schools) are reluctant to accept responsibility for students' classroom misbehaviour, according to the research, attributing fault to students and other non-school factors (Ho, 2004; Kulinna, 2007/2008; Miller, 2006). Such perceptions "have significant implications for teachers' perceptions of their own responsibility for students' performance as well as their subsequent behavior towards the students" (Ho, p. 376).

These research findings provide plausible explanations for why classroom site behaviour and management is seen as a source of tension for teachers. Teachers have been described as juggling competing attitudes regarding what they do (and what they should do) and being increasingly held publicly accountable and answerable (Beck & Young, 2005; Blanchard, 2003; Lingard, 2010). Friedman (2006) directly attributed high teacher burnout to

such educational trends as accountability and high-stakes testing, and further argued that students can recognise and be affected by teacher burnout.

2.2.2 Educational change at the school site.

2.2.2.1 *Educational change upon the classroom site.*

Educational change, including trends in Western countries' towards educational accountability, correlation of literacy standards with economic rationales, and high-stakes testing in schools, has been described as affecting Australian teachers' roles at the classroom level. Clarke (2001) noted that such changes have had impact "beyond the rights of individual teachers to ... the integrity of the teaching profession" (p. 46). Williamson and Myhill (2008) argued that a pervasive and growing market ideology in education and politicisation of education has contributed to increasing pressures on teachers' time, with growing work-related expectations upon teachers across Australia, including Tasmania.

Teacher deprofessionalisation and related issues were seen by Williamson and Myhill (2008) as consequences of these educational changes. A number of other writers endorse their views and recognise issues as including performance-based pay; trends towards uniformity and testing; and a divide between teachers' site-level focus on students and a system or bureaucratic focus on inputs and measurable outcomes (Beck & Young, 2005; Munt, 2004). Accountability requirements have been identified as affecting teachers at all stages in their careers, with some beginning teachers actively discouraged from exercising their professional judgement which "may

legitimate educational inequities by diminishing overt expression of control” (Achinstein, Ogawa, & Speiglmán, 2004, p. 593).

Educational change affects the classroom site particularly through expectations and pressures from the school hierarchy. In Olsen and Sexton’s (2009) school-based research, accountability and performance pressures are identified as originating from a broader government or social perception of crisis or issue, and culminating in pressures pushed onto the school and classroom level. Those authors described such tensions as being created and exacerbated “by centralizing and restricting the flow of information, by constricting control, by emphasizing routinized and simplified instructional/assessment practices, and by applying strong pressure for school personnel to conform” (Olsen & Sexton, p. 9). These analyses of hierarchical processes communicate a warning to all teachers in classrooms, not only those involved with testing required by education systems.

Teachers’ changing roles in this accountability-driven school context, according to Valli and Buese (2007), result in pressures for teachers at the classroom level to improve student literacy outcomes, as well as with “expanded responsibilities outside the classroom and intensified work within the classroom” (p. 523). Their research found that teachers’ negative opinions towards high-stakes tests were accompanied by teachers’ feelings of guilt, stemming from extolling the benefits of testing tasks of which they themselves were not convinced.

Educational accountability and high-stakes testing pressures have been further identified as contributing to teacher and student loss of motivation and related impacts on classroom control:

When teachers feel these administrative pressures, or when they believe that their students are extrinsically motivated or not motivated toward school, they are more likely to be controlling with students. It is possible that these conditions may directly affect teachers' behaviors or that they may undermine teachers' feelings or autonomy and motivation toward their own work, which in turn may lead them to be more controlling with their students. (Pelletier & Sharp, 2009, p. 180)

Finnigan and Gross (2007) argued that teachers' decreased motivation and other negative effects of accountability demands are exacerbated in schools or situations in which morale is also low. Further, "teachers responded less to the incentive (or threat) built into the policy and more to the value they placed on their professional status and the individual goals they had for students" (Finnigan & Gross, p. 624). Their work is consistent with research by Olsen and Sexton (2009) who pointed out that the classroom site, and working with students (rather than school administration), is the teacher's "refuge" or preferred consideration (p. 22). Similarly, Pelletier and Sharp (2009) stated that "the more teachers understood and agreed with the school mission and its associated goals, the less they suffered from emotional exhaustion and depersonalization, and the more personal accomplishment they felt" (p. 180).

Other research has reached similar findings. For example, Churchill, Williamson and Grady (1997) wrote that Australian teachers' cynicism about personal motivations behind changes in educational policy is intensified when

they perceive that changes are neither overly beneficial for teachers or students, nor as permanent. When teachers do accept the value and benefit of educational initiatives, on the other hand, claimed Muir, Beswick and Williamson (2010), teachers can be considered more likely to comply through their practice. Research by Brown (2004) concluded that “the success or failure of such policies may hang on the conceptions and meanings that teachers give to those policies” (p. 301).

Classroom change driven by policy demands can have both positive and negative results. Research notes that policy values interactive and well-paced teaching approaches (Cotton, 1995), but according to English, Hargreaves and Hislam (2002), the reality is that teachers often feel too pressured by competing academic demands to genuinely incorporate such methods into their classroom environment. Despite these findings by English, Hargreaves and Hislam, Warne (2006) suggested that classroom literacy teaching and learning can, in the face of increasing externally driven change, both cooperate with and minimise, or subvert and ignore, testing pressures.

Research results reflect a degree of ambivalence in the field; the full effects of these educational trends and changes remain obscured or unclear. Future research is indicated to identify and more fully consider impacts, at all levels of education and in government. Luna and Turner (2001) urged informative research to minimise negative classroom effects and build on teachers’ and students’ strengths, interests and progress.

2.2.2.2 *Senior school staff responses to educational change.*

The school impacts of accountability reforms and high-stakes literacy testing not only affect the individual teachers and students who face testing in the classroom. Senior school staff are also affected because they are a mediation point between system and site levels (Freebody & Wyatt-Smith, 2004), and are expected to oversee translation of educational changes into the school reality. As actors in this process, according to White-Smith and White (2009) and Wilson, Croxson and Atkinson (2006), principals too have particular experiences as a result of changes at the school level and their need to develop localised and considered responses to their situations.

Neoliberalism and market ideology in education, as already noted, have contributed to the popularity of school performance rankings. In the UK, the public demand for comparative school performance rankings, particularly between similar schools, has ensured senior school staff and principals' drive to achieve comparative school success (Wilson, Croxson, & Atkinson, 2006). The principal's role, already political and politicised, also involves expectations of mediating system-site translation issues (Crow & Weindling, 2010; White-Smith & White, 2009), such as having to either submit to or buffer schools from external accountability expectations (Rutledge, Harris, & Ingle, 2010). Research regarding school principals' roles demonstrates the importance of strong and positive principal leadership in mediating policy change at the school level. Principal leadership involves complex roles, and principals are expected to navigate varying demands exacerbated by the

already-present accountability and testing requirements (Crow & Weindling; Rutledge, Harris, & Ingle; White-Smith & White).

2.2.2.3 *Educational change above the classroom level.*

The current education context for Australian schools places contrasting demands on those who work at the school site. Florian and Rouse (2001) highlighted the often-contradictory expectation upon schools, particularly government-run schools, to engage in inclusive practices, while simultaneously demonstrating constant student academic improvement. Schools play a crucial role in relation to student academic success (or failure), which has translated into policy and curricula demands for targeted student improvement (Blanchard, 2003). In many countries, including Australia, what has been termed an ‘achievement gap’ is particularly evident across lines of race and ethnicity (Nguyen, 2010; Sherman & Grogan, 2003). Stakes are high, both at the system and the (school) site levels, to not only be seen to be addressing student non- or under-achievement, but to in fact work towards solutions and improvements. It is those actors at the school and classroom sites, however, whose work towards these goals is most vigorously scrutinised and held accountable.

School-level education does not occur in a contextual vacuum, and literacy teaching cannot be reduced to a prescription; much work goes into teaching and assessing student literacy at the individual and class levels, a “mix [of] psychology, history, literature, politics, sociology, linguistics, economics, art, science, philosophy, poetics and aesthetics with passion and dispassion,

with pragmatism and vision” (Boomer, 1998, p. 20). In contrast, however, education policy pushes uniformity and standardisation, economic ideologies and increasing management, which negates the importance of teachers’ and other school actors’ expertise and knowledge (Grant, 2009; Smith, Edwards-Groves, & Kemmis, 2010). Brooks, Maxcy and Nguyen (2010) reflected that “educational prospects for students of differing backgrounds and characteristics – ethnic, religious, cultural, economic, linguistic – remain unsettled and unsettling” (p. 3), and argued that changes in educational policy and system demands upon the site level should consider the value within difference. Those already working hard at the site level, those authors say, will likely struggle to implement imposed context-inconsiderate mandates (Brooks, Maxcy, & Nguyen).

A range of policy implementation issues is raised in the literature. For example, for Busher (2006), policy should reflect site realities and needs, particularly those of students’ effects upon the classroom site, involvement in decisions to directly affect them, and enthusiasm for learning or other classroom tasks. Such moves, he urged, are the only way of ensuring student schooling success (also Brookhart & Bronowicz, 2003; Dutfield, Allan, Turner, & Morris, 2000; Whitty & Wisby, 2007; Zion, 2009). Given that research has suggested that students absorb media messages about market and individualist ideology (Ochoa & Pineda, 2008), schools could be encouraged to incorporate these proposals into implementation strategies. Apple and Beane (2007; Beane & Apple, 2007) strongly argued the importance of positive school-wide

responses to current trends in order to empower all school actors and engender education institutions that could provoke positive social change.

2.2.3 Teachers' mediation of literacy assessment.

Literacy teaching and assessment has undergone changes in recent decades, as previously outlined, and literacy teachers have experienced conflicting demands (Blanchard, 2003; Churchill, Williamson, & Grady, 1997; Lingard, 2009, 2010). In an educational context that mandates compliance with standardised literacy testing, Australian literacy teachers are also nonetheless required to conduct class-level assessments in order to gain formative and summative evaluations of student achievement. Teacher assessments are expected to contribute information at the individual student level through to the school-wide knowledge base regarding effectiveness and appropriateness of literacy syllabi and related programmes (Wyatt-Smith, 2008; Wyatt-Smith & Campbell, 2004).

The literature reveals how assessment activities of literacy teachers have been affected in a variety of ways by this changing context of literacy education. Teachers' assessment judgements occupy an uncertain position with accountability and high-stakes testing. Harlen (2005) noted continuing discussion about the extent to which teachers' assessments, used to promote learning motivation, actually differ at times from realised student achievement on high-stakes standardised tests. Reviewing a number of studies on this question, Harlen wrote,

[T]here was evidence of low reliability and bias in teachers' judgements made in certain circumstances. However, this has

to be considered against the low validity and lower than generally assumed reliability of external tests. (p. 245)

Teachers use their own 'indexes' for making assessment judgements (Wyatt-Smith, Castleton, Freebody, & Cooksey, 2003), and the literature acknowledges that some 'unreliability' might be evident across teachers' literacy assessments. Teachers are expected – required – to incorporate set criteria for assessment benchmarks, against which their assessment judgements are to be validated. Such assessment criteria and standards provide an explication of what can be considered quality in student work and in teacher assessment of student work against which teachers must consider their own assessment activities (Cumming, Wyatt-Smith, Elkins, & Neville, 2006; Wyatt-Smith & Klenowski, 2008). These criteria and standards must be clear for teachers and, arguably, relevant at the teachers' site level. Brown (2004) warned, "new assessment policy, tool, or practice, whether at the national or local school level, needs to take account of the complex structure of teachers' conceptions of assessment to ensure success" (p. 314).

Teachers' assessment judgements must "speak simultaneously to potentially conflicting demands" (Cooksey, Freebody, & Wyatt-Smith, 2007, p. 403) across system and site levels. Considered in this light, high stakes literacy tests too could be seen as subject to conflicting demands. In a similar way that teachers can be positioned in this assessment environment, the literacy tests themselves could potentially be treated as tools to meet both system and site expectations and needs.

New conceptualisations of multiple forms and expressions of literacy have meant changes in approaches to assessment, including a popularity among

literacy teachers for utilising broader, holistic measurements instead of “one-off, single testing instruments” (Fehring, 2005, p. 95). The literature review recognises a lingering gap in understanding how Australian teachers reach their assessment ratings and decisions. Education research is now beginning to explore this aspect of teaching (Castleton, Wyatt-Smith, Cooksey, & Freebody, 2003; Cooksey, Freebody, & Wyatt-Smith, 2007; Wyatt-Smith, 1999; Wyatt-Smith & Castleton, 2004). Research points to a number of factors, at times seen as problematics, affecting teachers’ assessment practices including community context, teacher experience, moderation practices, assessment criteria and standards, observations of the student and knowledge of pedagogy. Researchers have noted too that, at the point of assigning student work, some of these criteria may not be accessible for teachers (Castleton, Wyatt-Smith, Cooksey, & Freebody; Wyatt-Smith, Castleton, Freebody, & Cooksey).

Questions about the reliability, validity and accuracy of teachers’ student assessments garner attention. Castleton, Wyatt-Smith, Cooksey and Freebody (2003) remind that a lack of public clarity about the often implicit processes through which assessment takes place is not intentional, but is representative of “the complex and dynamic nature of acts of judgement, and how teachers themselves do not intuitively map acts of judgement as they occur” (p. 41). However, those authors argued that an understanding of the judgements teachers make in engaging in literacy assessment is necessary, given an education context that demands transparency in all areas.

Australian teachers currently operate in an environment with increasing expectations about transparency, despite significant contestation about the

work itself. As already noted, the importance placed on assessments and evaluations of student learning is being particularly driven by external forces (Au, 2007, 2008). Numerous researchers (for instance, Beck, 2008; Beck & Young, 2005; Lingard, 2009, 2010) have described literacy as a site of conflict where navigating contrasting understandings and expectations of validity within the site (school) and system levels can create intense pressure on the teachers. As Sim (2006) argued, “the increase in accountability of schools undertaking [literacy] programs has provided a prompt to examine the literacy practices within the whole school and the role of each teacher within these practices” (p. 240).

Although teachers hold their own beliefs and opinions about what constitutes student achievement in literacy learning (Wyatt-Smith & Castleton, 2004), they are nevertheless required by the broader education system to be dedicated to assessment activities and processes deemed valid by that system (Freebody & Wyatt-Smith, 2004). Teachers also face requirements for their participation in assessment activities and processes that are valid and relevant for their particular and unique school site (Freebody & Wyatt-Smith). Research by Cooksey, Freebody and Wyatt-Smith (2007) showed that a major focus of studies on teachers’ assessments is “on the quantification of student outcomes or on the descriptive study of assessment processes” (p. 402).

The literature on literacy teachers’ assessment activities supports the understanding of assessment as a complex task (Cooksey, Freebody, & Wyatt-Smith, 2007), through which teachers must draw on their expectations of students as well as their knowledge of curricula and other school needs in order

to be site-valid (Freebody & Wyatt-Smith, 2004). Considering the resource represented by teachers' assessments, researchers such as Brown, McCallum, Taggart and Gipps (1997) and Zyngier (2009) have urged that teachers' assessments be used to contribute to the broader corpus of data on students' literacy progress and standards. It is in this frame of recognising teachers' expertise in formulating and utilising assessments that researchers such as Sloane and Kelly (2003) argued the merits for teacher involvement in test design and content.

2.3 Summary of Literature Review

This literature review has concentrated on those works broadly related to the dual research focus: standardised literacy testing itself, and the classroom site of such tests. A range of issues affecting educators, students, parents, government and the general public have been identified and examined to build on the researcher's previous understanding from professional and practice experience. This research has recognised the significance of the broader social and environmental context affecting specifics of the classroom site and a critical overview of these connections emphasised in the literature has been presented. From reviewing key writings in the literature, a picture forms of composite layers of expectations and demands on literacy teachers, inevitably creating 'high stakes' status for these assessment practices in today's education context.

The literature demonstrates that a consensus has yet to be reached concerning the value and validity of standardised literacy testing, although research identifies that such tests serve information, accountability, monitoring

and comparison purposes. The literature further demonstrates that patterns of teacher-student classroom behaviour also depend on context, circumstance and relationships, indicating that classrooms and their actors can share differences and similarities. One key message strongly echoed across the research regarding both research questions is the importance of system recognition, through educational policy, of the complexities at the (school) site and the importance of policy reflection of realistic site possibilities and context.

Chapter Three: Methodology

Introduction

This chapter describes and elaborates the methodological approach adopted for this research study. This qualitative case study is framed by the research questions and theoretical principles, and undertaken through a combined constructivist grounded theory (Charmaz, 2000, 2001, 2005, 2006; Charmaz & Mitchell, 2002) and discourse analysis (Carabine, 2001; Fairclough, 2001; Gee, 1999, 2005, 2011; Taylor, 2001a, 2001b) approach.

This chapter is structured in seven sections. The first section, 3.1 Research Questions, presents and overviews the two overarching research questions that are addressed through this research study. The second section, 3.2 Selection of the Research Approach, briefly outlines the qualitative, case study approach employed within the study. The third section, 3.3 Theoretical Principles, presents the theories influencing the researcher's approach to the analysis, particularly the researcher's understanding of discourse and employment of the selected discourse analysis approach.

The fourth section, 3.4 Development Phase of the Research, relates the steps undertaken to collect and generate the data, including University of Tasmania and Department of Education Ethics approvals, site and participant selection, and information letters and consent forms for research participants. This section, too, clarifies the type of sample used in the study, and reiterates the number of participants, composition of observed class groups, and the size and description of the school site. This section also outlines the development

of the research instruments: questionnaires, open-ended interview schedules, and non-participant researcher observation. In this section, a matrix of timelines for the research approval, data collection and analysis also communicates those research stages completed and the sequence of activities.

The fifth section, 3.5 Data Generation and Collection, introduces the forms of data generated and collected for analysis: testing session transcripts and field notes recorded during periods of non-participant researcher observation of testing sessions, teacher and student questionnaires, and teacher and principal interviews. This section discusses the collection and generation of data by data type, overviewing how and when each type was collected, and identifying how the forms of data were used to address the research questions.

The sixth section, 3.6 Approaches to Data Analysis, overviews the constructivist grounded theory (Charmaz, 2000, 2001, 2005, 2006; Charmaz & Mitchell, 2002) and discourse analysis (Carabine, 2001; Fairclough, 2001; Gee, 1999, 2005, 2011; Taylor, 2001a, 2001b) employed in the data analysis. The final section, 3.7 Structure of Analysis Chapters, very briefly reiterates the analysis stages, the results of which are presented in chapters four to six.

3.1 Research Questions

This research project is undertaken within the context of standardised literacy testing in a Tasmanian secondary-level State school. In order to examine issues, approaches and effects of standardised literacy testing, the analysis results chapters (chapters four to six) will report the data then contextualise the work. Understanding the context and elements of the tests' administration is

necessary to begin an informed examination of the effects of test administration upon all involved in the processes. In brief, this context as outlined in Chapter One includes: the ways that standardised literacy testing can be seen to take place in classrooms, and the justifications teachers give for their test practices; the educational and political outcomes and uses of standardised literacy testing, and the testing participants' perceived roles in standardised literacy testing practices and processes; the ways that classroom test actors (in this case, teachers and students) interact through standardised literacy test sessions, and ways that this interaction affects the testing; and test actors' perceptions of the benefits and uses made of standardised literacy testing and test data, and for whom and why teachers consider test data to be beneficial and useful. The context and research topic are examined through the following two related research questions, namely:

Research Question 1: In what ways do teachers account for the value and validity of standardised literacy testing?

This research question frames the investigation as one centred on a consideration of teacher participants' understandings and opinions of specific aspects of standardised literacy testing processes and practices. 'Value' here refers to the values that are made of tests and test data, such as the uses considered appropriate and the reasons for these uses. 'Validity' refers to the accuracy of test data and results as representations of student literacy levels, for instance compared with teachers' assessments, rather than a technical term denoting the particular utilised statistical or computation methods. Of interest

here are teacher participants' ways of accounting for tensions between the participants' administrative role of 'tester' and that of 'teacher', and of tensions between on the one hand their interpretation of perceived expectations and demands of 'testing', and on the other hand behaviours and responses they considered necessary due to the needs of the student group. This research question acknowledges teachers as more than de-professionalised instruments employed in the gathering of student literacy data.

Research Question 2: What patterns of action and interaction characterise the classroom site of standardised literacy testing?

With this research question, the direction of the study moves to the classroom site of Tasmanian standardised literacy testing. This research question highlights as relevant for research the practices and processes engaged in by teachers administering the tests, as well as the ways in which classroom test actors engage with the tasks. 'Action' here refers to the ways in which testing administration is executed: how can teacher administration of standardised literacy testing be understood and described; what kind of an experience or enterprise is teacher administration of testing? 'Interaction' expands the focus to deliberate or unintentional consequences of test administration, examining effects of test administration upon test actors (a term which, to reiterate, means in this research all those who are directly involved in the school site of standardised literacy testing: teachers and students, and school principals), and test-related interplay between test actors.

3.2 Selection of the Research Approach

3.2.1 Qualitative research.

This research study fits within a broad qualitative research paradigm.

Qualitative research has played a critical role in the development of educational theory and practice; qualitative research has contributed to knowledge in the areas of “educational inequality, socialization and identity formation, school organization, and educational policy” (Riehl, 2001, p. 115), and qualitative researchers “generat[e] theoretically rich examinations of schools as organizations, addressing issues of culture and education” (Riehl, p. 115), developing broader social analyses of schooling and society, and incorporating perspectives on learning as situated, sociocultural activity to research concerning schools as contexts for teaching and learning.

This study incorporates a number of broad identifying features of qualitative research: the inquiry takes place in a ‘natural’ setting (in this case, the school site) and is interpretive in character. The process adopted is inductive and flexible, with the researcher assuming the role of data collector and analysis instrument throughout the process, and valuing ‘thick’ description through detail from the data (Imel, Kerka, & Wonacott, 2002). ‘Thick’ description is defined as the theoretical interpretation of data enriched through careful, deep analysis of and immersion in all the collected data. ‘Thick’ description is the point at which the analysis presents grounded, illuminating examination of the data.

This is a qualitative research study, shown in its approach and through the data gathering and analysis phases; in addition, it employs elements of a quantitative approach, of frequency counts, calculations and comparison

necessary to tally, compare and combine the constructed codes and categories. The constructivist grounded theory coding was not an attempt to “reduce the [participants’] responses to quantitative categories” (Smith, 1999, p. 9), but rather an attempt to organise themes and patterns to aid the construction of thematic areas and discourses. Arguably, this (although very minimal) use of a mixed-methods approach strengthens the comparative capacity across data, without undermining the strengths brought to the research by the qualitative interpretation and analysis.

3.2.2 Case study.

This investigation utilises a case study method (Freebody, 2004; Isaac & Michael, 1995). Being a case study, it examines its topic, which is the context, administration and effects of Tasmanian secondary-level State-school standardised literacy testing, through the perspective or lens of a particular setting and group of social actors, a setting that constitutes “a social unit in its own right ... a holistic entity” (Payne & Payne, 2004, p. 31). This setting is a particular school and classroom site.

Case study research is valuable in education in that it is flexible and naturalistic, and allows research to be grounded in education settings. Case studies illuminate the ‘real world’ experiences of the actors in particular instances, and allows for examples of phenomena as they occur or have occurred at one point in time: case study research “provides a unique example of real people in real situations, enabling readers to understand ideas more clearly than simply by presenting them with abstract theories or principles”

(Cohen, Manion, & Morrison, 2000, p. 181). Classroom and school sites are complex sites for research, and the case study method is ideal for such fieldwork research as it can consider a smaller instance and illuminate what can be missed when the data are much broader in scope (Burns, 2000).

The theoretical stance that informs the research, such as the researcher's poststructuralist leaning and the understanding that no research occurs in a value vacuum, shaped this case study approach. As such, the claim that "no value stance is assumed" in case study research (Anderson, 2004, p. 152) has resonance in this case study. A value stance is assumed as important throughout the process beginning with initial selection of the research topic and including the researcher's identification of classroom administration of standardised literacy testing as a source of worthy data together with consideration of classroom actors' testing experiences as worthy of focus.

3.3 Theoretical Principles Underpinning the Research

3.3.1 Postmodernist and poststructuralist theory.

This study is influenced by tenets and theories of postmodernism and poststructuralism, which although closely related and often confused, are distinct (Wright, 2003). Lather (1991) posited that postmodernism refers to the "larger cultural shifts of a post-industrial, post-colonial era" (p. 4), and poststructuralism to the "working out of those shifts within the arenas of academic theory" (Lather, p. 4). Poststructuralist theory examines and highlights "multiple realities, shifting subject positions and incomplete truths" (Carroll, 2002, p. 1). This theoretical positioning "does not recognize [sic] the existence of unambiguous objectivity; the way individuals construct the

meanings of phenomena is paramount ... qualitative specialists (especially postmodernists) are more likely to acknowledge multiple interpretations” (Whipp, 1998, p. 52).

No single postmodern or poststructural theory or approach exists, but the basic premise underlying them is that the “technologies of communication profoundly shape human experience” (Lather, 1991, p. 21). In this research, such an argument means a belief in the value of opening up participants’ voices, to disrupt a dominant, hegemonic ‘reality’ of the classroom standardised literacy testing experience. Within education, this framework encourages recognition of student, school and contextual difference, just as this framework values an acknowledgement of difference and diversity (O’Farrell, 1999). Participants’ ‘voices’ are incorporated in the analysis results chapters (chapters four to six) to identify the ways they construct their own meanings of standardised literacy testing. This incorporation of participants’ voices and consideration of practices they have normalised or problematised aids deeper examination and a disruption of what participants indicate are normative representations and expectations within their school site of testing.

This researcher considers that postmodernist and poststructuralist theories of language, subjectivity, social processes and institutions are all relevant to understanding “existing power relations and to identify areas and strategies for change” (Humphries, Mertens, & Truman, 2000, p. 11). Individuals are “shaped” by the “social institutions and practices” which produce social “meanings”; however, it is the individuals who are “agents of change” within these institutions, practices and meanings, their change either

serving “hegemonic interests” or challenging “existing power relations” (Weedon, 1987, p. 25). An interest in hegemonies and norms, and the ways that they are constructed, perpetuated and challenged, supports and informs the research examination of the data for relevant patterns. The employment of constructivist grounded theory coding and discourse analysis approached synthesise with and reflect a sympathy for postmodernist, poststructuralist theories through the problematising of assumed objective, positivist research.

This work selectively utilises postmodernist and poststructuralist theory. The chief aspects of these related theories are their challenging of hegemonic, dominant voices and a valuing of various interpretations. Through this perspective, the research undertaken focuses on the interpretations and perspectives of those who directly experience the testing phenomenon at the organisation, administration and follow up stages; indeed, the site level experiences, rather than the normalising and hegemonic system level stance on the phenomenon. A further note on these theories is that, for purposes of consistency, this work will refer to these related positionings and their value for the study by the single term poststructuralism.

3.3.2 Discourse.

Considerable debate surrounds what constitutes discourse (Edley, 2001; Stubbs, 1998; Taylor, 2001a, 2001b), and the term and approach are used differently in different fields of research (Phillips & Jørgensen, 2002; Taylor). This research supports the position that, as researchers’ decisions reflect their beliefs and values (Naples, 2003), so the concept of discourse that a researcher

follows is a value-based choice. The concept of discourse used in this study reflects the researcher's understanding regarding knowledge and power structures, the construction and perpetuation of social realities, non-neutrality of language, and value of a frame of reference with which to examine and disrupt hegemonies.

The concept of discourse adopted in this study is also informed by Gee's (1999, 2005, 2011) theory of 'big D' and 'little d' discourses. Gee defined 'little d' discourses as smaller, linguistic or language-based discourses, compared with 'big D' Discourses, which are "language *plus* 'other stuff'" (2005, p. 26). Gee (2011) used 'big D' Discourse to denote "ways of combining and integrating language, actions, interactions, ways of thinking, believing, valuing and using various symbols, tools, and objects to enact a particular sort of socially recognizable [sic] identity" (p. 201). In this study, 'little d' discourses are the smaller wording and language choices made by participants, noted through the close reading of the transcripts and incorporated into the analysis through the coding, while the 'big D' Discourses are larger-scale meanings shown through language choices, patterns of behaviour, meaning-making, and social identities. However, as this study is chiefly focused on the latter type, 'big D' Discourses will not always be capitalised, and will also be termed 'dominant discourses'.

Discourse does not merely communicate norms, but establishes them (Carabine, 2001). Similarly, discourse "does not just contribute to the shaping and reshaping of social structures but also reflects them" (Phillips & Jørgensen, 2002, p. 61). Further, discourse both constructs and is constructed by social

patterns and practices; “discourse and practice are inextricably bound up with one another” (Edley, 2001, p. 192). As such, discourse is connected with socially constituted and constitutive ‘truths’, norms, accepted knowledge and power, and the ways that we consider, construct and experience our ‘realities’ are in no way value-free.

Such a conceptualisation of discourse understands power as inherent within discursive operations, and understands power and discourse to be manifested and problematised with discursive actors’ access to or use of knowledge. Here, knowledge is the location where discourse becomes normalised or challenged, and where power can be countered or maintained. An examination of the relationship of power, discourses and knowledge within the school site of standardised literacy testing can, to use the words of Carabine, help to clarify the “nature of power/knowledge in modern society” (Carabine, 2001, p. 276). Discourses occur “within a shifting and dynamic social context in which the existence of multiple sets of power relations are inevitable”; as a consequence, “[d]iscourse and politics, knowledge and power [are] part of an indissoluble couplet” (Apple, 1991, p. vii).

The discourse analysis approach adopted in this research facilitates an examination of “power and resistance, contests and struggles” (Taylor, 2001a, p. 9), and is undertaken with an appreciation of the power of discourses in people’s different social realities and positions. The discourse analysis approach employed in this research seeks to examine and deconstruct power relations and structures pertinent to school-site standardised literacy testing that

are indicative in the data, in particular those that test actors (and others outside the school site) knowingly or unknowingly perpetuate or challenge.

Hegemony “emphasizes [sic] forms of power which depend upon consent rather than coercion. ... Hegemonic struggle penetrates all domains of social life, cultural as well as economic and political, and hegemonies are sustained ideologically, in the ‘common sense’ assumptions of everyday life” (Fairclough, 2001, p. 232). With its interest in power, knowledge and discourse, this study also examines hegemonies evident in participants’ test experiences. Issues of hegemony are relevant to this study, given issues of ‘stakes’ attached to standardised literacy testing already discussed in chapters one and two. Discourse, as understood in this research, is permeated by and affects hegemonic or dominant practices and forms of ‘truth’ and knowledge, which can then be questioned alongside or within the discourses themselves.

Language, as communication of roles and ideas, social semiosis and practice, is the site of the defining and contesting of social organisation and their consequences, as well as the site through which subjectivity is constructed (Weedon, 1987). As “language available to people enables and constrains not only their expression of certain ideas but also what they do” (Taylor, 2001a, p. 9), it “constructs the individual’s subjectivity in ways which are socially specific” (Weedon, p. 21).

Language is used to further particular discourses, such as expanding normative patterns and identities, establishing forms of knowledge and ‘truth’ (Fairclough, 2001; Phillips & Jørgensen, 2002). The understanding of discourse within which this research is couched problematises assumptions of

“the neutrality and reflectiveness of language” (Taylor, 2001b, p. 316) and instead considers that “language connects with the social through being the primary domain of ideology, and through being both a site of, and a stake in, struggles for power” (Fairclough, 1989, p. 15).

Language provides a framework through which people can understand, maintain or challenge hegemonies and norms, whether, for instance, social, cultural, legal or ethical. As such, language in all forms both reflects and is dialectically involved with discourse. The research participants and researcher are located within the struggles and constraints of discursive patterns, norms, and within a context created by and through discourse and language. Actors’ discursive positions and reactions to the various hegemonies and norms within which they operate are considered to be powerful because this researcher holds that “some ways of understanding the world can become culturally dominant or hegemonic ... they can assume the status of facts, taken for granted as true or accurate descriptions of the world” (Edley, 2001, p. 190).

This understanding of discourse values a questioning of discourses’ “wider implications, such as the identities they make available and the constraints which they set up” (Taylor, 2001a, p. 9). As argued by Phillips and Jørgensen (2002), “subjects do not become interpellated [sic] in just one subject position: different discourses give the subject different, and possibly contradictory, positions from which to speak” (p. 17); subjectivity “is produced in a whole range of discursive practices ... the meanings of which are a constant site of struggle over power” (Weedon, p. 21). These discourses and

discursive practices and positions need to be analysed with regard to the contexts of which they are a part.

3.3.3 Subjectivity and researcher reflexivity.

The relevance of subjectivity within a poststructuralist-influenced discourse analysis study lies in the value of questioning human understandings of knowledge and subjectivity, integral to a discourse analysis orientation. The appreciation of subjectivity is central within a study of discourses that questions “what it means to be human, what counts as ‘real’ and what the ‘social’ is” (Taylor, 2001a, p. 5). A belief in the constructed and constituted nature of meanings, knowledge and discourse is influential upon this research. As a result of these powerfully constituted and constitutive discursive processes, subjectivities become our way of recognising and interacting with the world, through its norms and hegemonies. Subjectivity and interaction enable knowledge to become accepted and normalised in society (Phillips & Jørgensen, 2002).

Everything is open to such multiple and individual interpretations and, as posited by Apple (1991), “Reality it seems is a text, subject to multiple interpretations, multiple readings, multiple uses” (p. vii). The best the researcher can do is offer an interpretation of these alternative perspectives and subjectivities, though; as argued by Geertz (1973), “What we call our data are really our own constructions of other people’s constructions of what they and their compatriots are up to” (p. 9), and Taylor (2001a) too reminded, “the discourse analyst is not outside these struggles and constraints but is one such

user within them” (p. 10). Lather (1991) claimed to aim for “multi-voiced” (p. 9) analysis, which incorporates participants’ voices to decentre the author and acknowledge the author’s location within discourses, and here the understanding of discourse meets the poststructuralist appreciation.

The connectivity between this idea of subjectivity with poststructuralism means that the researcher must consider their influence upon the research and potential location within the discursive struggles being studied. Reflexivity in research is considered in this research to mean more than merely “situating where one is coming from” (Woodward, 2000, p. 43), but rather “the continual assessment of the contribution of one’s knowledge to others, as well as the questions we have asked; the way we locate ourselves within our questions and the purpose of our work” (p. 44). This means “remaining sensitive to the perspectives of others and how we interact with them [and] a continued self-awareness about the ongoing relationship between a researcher and informants” (Naples, 2003, p. 41), to challenging “the authority of the author and/or of the power difference in the field” (Wasserfall, 1997, pp. 151-152, quoted in Naples, p. 42). In this way, researcher reflexivity is an extension in the research process of the researcher’s perspective on subjectivity and multiple realities. This researcher considers that the researcher subjectively influences all stages of the research, and, although not attempting to negate these influences, the researcher can critically question their involvement and role within power structures in the research, to produce more honest and critical research. Indeed, although both researchers and participants bring contextual knowledge and subjective interpretations to research, it is the

former who “are obligated to be reflexive about what we bring to the scene, what we see, and how we see it” (Charmaz, 2006, p. 15).

Researcher reflexivity can represent a “profound, personal transformation in one’s consciousness of oneself, of others and one’s relationship in the social world of which one is a part” (Applebaum, 2001, p. 417), important because a genuine “critical self-examination (self-location) can help unmask hitherto hidden privilege” (Johnston, 2001, p. 424). This researcher is located within the studied discursive struggles, as both a teacher of secondary-level English/Literacy and as a member of the media-consuming public. These two positions play different roles within the processes by which hegemonic expectations and messages are constituted, created, consumed, accepted and problematised, about literacy standards, testing, and teacher roles within standardised literacy testing.

The theory that discourses and social phenomena are not value-free but are socially and politically constituted and constitutive is sympathetic to a poststructuralist-influenced research approach that incorporates the principles and practices of subjectivity and reflexivity. The idea of subjectivity influential upon this research problematises ideas of ‘truth’ and ‘reality’, and enables inclusion and exploration of different interpretations and discursive positionings, as “Our knowledge of the world should not be treated as objective truth” (Phillips & Jørgensen, 2002, p. 5). This research is not undertaken in an attempt to identify the ‘truth’ or the actors’ intent; instead, the focus is on the discourses that the analysis makes apparent, how they appear to be constructed and framed, and their possible effects (Phillips & Jørgensen). The influence of

poststructuralism on this study directs the researcher and the research activity towards a sympathy for differing subjectivity and ‘realities’.

3.3.4 Problematising validity, reliability and credibility.

Influenced as it is by poststructuralism and subjectivity, this work is not intended to produce objective, empirical results, but instead to explore discourses within participants’ perceptions of their experiences. Instead of aiming for “consistency in measurement”, truth of interpretations, and “generalizability [sic] of the obtained results” (Anderson, 2004, p. 12), this researcher accepts that the work produced “is *one* reading of the texts”, and “there will be numerous other readings possible” (Duncan, 1996, p. 161, quoted in Neuendorf, 2002, p. 6), and that the ‘reading’ produced will reflect researcher and participant subjectivities.

The participants’ experiences, highlighted by the questions asked of participants, the methods of analysis and influence of poststructuralist theory, and the data, will be explored through this research. As the research is concerned with the focalising and exploring of participants’ experiences, the replication of findings in other research is limited, as different participants will have different interpretations and contextual understandings. As such, this researcher problematises concepts of validity, reliability, credibility and replicability, and holds that the produced work is valuable and important due to the grounding of codes in the data, the contextualising of themes and discourses in other research and literature, and the contributions made to the theoretical, methodological and knowledge fields.

The reaching of similar or identical findings in other studies might or might not be possible, but this researcher does not hold this to be the most important criteria for this research. Rather, the research and its findings contribute to avenues through which the experiences of participants are valued and considered (Apple, 1991). The stance adopted in this research regarding the importance of questioning claims of “value neutrality” (Naples, 2003, p. 52) challenges the “relations of ruling embedded in the production of knowledge in the academy” (Naples, p. 52) and highlights participants’ experiences. The findings from this research will be contextualised with an examination of these research findings against those from other research and literature. In addition, the constructivist grounded theory and discourse analysis are grounded in the data due to the inductive theorising processes. In this way, the combined approach can serve a validation purpose if necessary, and provide substantiation for analysis, in acknowledgement of or to meet that intent that the academic research context itself establishes particular requirements.

Similar to grounded theory research projects conducted by Corbin and Strauss (1990), this work aims to explain and describe, change and take a stance in relation to determinism, while also attempting to avoid the “dangers that lie in [the] positivistic connotations” behind many of the necessary elements of verification, consistency and relevance (pp. 4-5). The aims of this study mirror those articulated by Corbin and Strauss, being to “uncover relevant conditions”, “determine how the actors respond to changing

conditions and to the consequences of their actions” and, importantly, to “catch” the “interplay” (Corbin & Strauss, p. 5).

In the process of fulfilling these aims, this research project attempts to work towards the “plausibility and value of the theory itself or ... of its modest theoretical formulations”, the “adequacy” of the research process and the “grounding” of the findings (Corbin & Strauss, 1990, p. 16). The analysis will highlight what the data indicate to be discursive struggle by examining the data and interpreting the themes through a triangulation of data analysis approaches (constructivist grounded theory and discourse analyses) and data types (questionnaires, interviews, session transcripts and field notes, collected from within the fieldwork setting and compared with collected research and literature). Triangulation, to “establish converging lines of evidence to make your findings as robust as possible” (Yin, 2006, p. 115), is not intended to show the validity of findings, but to generate richer themes and perspectives on the phenomena.

3.4 Development Phase of the Research

This research study involved a number of steps prior to and within the generation and collection of the documentary and fieldwork data. The Tables below show the order in which the data were generated and collected, following which the steps will be explained.

Table 3-1

Timeline of data collection

[illegible]

Table 3-2

Steps taken within each stage of data collection

Activity	Details of activity
Preliminary contact with Principal re site selection	Contact made with principal, gained initial research approval, sent information letter and P's permission letter Sent P teacher and student questionnaires, and teacher interviews
Principal interviews	Initial and second P pre-interviews P, Interview 1
TT interviews	TT, Interviews 1-3
Observed lesson	Year Seven, Humanities
Observed testing sessions	Year Nine test sessions: Writing (over two lesson sessions) Reading/Comprehension Spelling Year Seven test sessions: Reading/Comprehension Writing Spelling
Student questionnaires	Year Nine: end of Testing Session 3 Year Seven: end of Testing Session 6
Research information (interviews) provided for participants	Information letter, consent form and final interview schedules provided to P and TT
T1 and T2 interviews	T1, Interview 1 T2, Interview 1

3.4.1 Initial research approval: Ethics.

Before selecting and contacting the fieldwork location of one Tasmanian secondary-level State school, the research study required approval from the University of Tasmania's Faculty of Education ethics board and the University of Tasmania Human Research Ethics Committee. Approval was also sought from the Tasmanian Department of Education, because of the intention to

undertake research in a government school. Approval documentation is attached as Appendix A.

3.4.2 Site selection.

Following University and Department of Education research approval, the research site could be selected and approached. As the intended school site for the research project was a secondary-level State school, the researcher began with a list of such schools in the broad geographical area of the city of Hobart. In order to gain more broadly relevant fieldwork data and results, single-sex schools were removed from the list of options, and the researcher contacted the principals of two of the remaining schools, from different areas in the wider city zone. Initially, email communication ascertained the principals' levels of interest in the research, and the one principal who agreed was then contacted more formally. Letters regarding the project were endorsed by those involved in the research: the doctoral candidate, research supervisors, and the University of Tasmania Faculty of Education. This formal, written communication included information on the purpose of the research, the participants needed (teachers of English/Literacy who would be expected to administer the upcoming standardised literacy tests), what the participants would need to do to take part in the research project, the time frame for the study, and the assurance of participant anonymity.

The site was selected with the decision already made to use it as a small case study site. The study was not intended to capture an exhaustive picture of participating teachers' experiences with standardised literacy testing, but begin

to explore some of the ways that the phenomena of standardised literacy testing took place at one school. As such, only a small number of teacher participants and the one school principal would be needed to generate and collect the necessary data. This small type of sample makes data easier to generate, collect and analyse, but makes findings less generalisable and transferable. This qualification is relevant to consideration of the findings from this study.

The school site at which the data were collected was in an outer suburb of Hobart, capital city of Tasmania. The school teaches year levels seven to ten. The school was a secondary-level public school, meaning the school was funded entirely by Federal and State governments, and the school was accountable for its operations and curricula decisions to government. During the school year when the data were collected, 295 students attended the school, and staff included 26 teaching and 26 non-teaching staff members, not all full-time employed. Along the Educational Needs Index (ENI), the selected school site has a rating of 66 out of a possible 120, placing it only slightly towards the more needs end of the spectrum. According to the Tasmanian Department of Education, this “ENI is used to allocate resources under a number of programs [sic], including teacher staffing and the General Support and Maintenance Allocations in the [Schools Resource Package]” (Department of Education, 2010). The school’s ‘feeder’ primary schools were positioned further towards this needier end of the ENI spectrum.

3.4.3 Sample.

Lankshear and Knobel (2004) discussed sampling methods with regard to quantitative, rather than qualitative, research. “In quantitative research, investigators often are interested in drawing inferences on ‘generalizing’ [sic] from a sample of observations to a broader population” (Lankshear & Knobel, p. 147). Being able to demonstrate the generalisability of research findings is less relevant to this particular research than to research produced within a quantitative paradigm that chiefly aims for objectivity and truth. As such, an extensive justification of the sampling approach employed in the data collection stage will not be provided, as generalisability and validity are not claimed by this research. However, the sampling that was used can be described with reference to both purposive and convenience sampling (Lankshear & Knobel).

The school site was selected through a deliberate decision regarding the broad geographical location of the school (purposive), but was one of two schools approached and the only school that agreed (convenience). The selection of the participants, similarly, was both purposive and convenient, as only English/Literacy teachers who would be involved in the standardised literacy test administration were to be involved (purposive), and the teachers available were included as participants (convenience).

Any issues inherent in purposive and convenience sampling are not considered by this researcher to undermine the value of the produced research. This research aims not to examine and provide a single truth that can be verified at different research sites or through different research methods as with

a positivist research paradigm, but instead to analyse the discursive patterns and tensions highlighted through the collected data in an attempt to explore representations of the issues and their effects (Lankshear & Knobel, 2004).

3.4.4 Information letters and consent forms.

Formal approval having been granted from the school site meant information letters and consent forms could be sent to the principal for distribution to teaching staff who would be involved with the standardised literacy testing at that school. From this process, the principal and three teachers agreed to participate in the research. The signed consent forms were returned to the researcher, the researcher met individually with the principal and teacher participants to discuss and clarify the research project, and the data collection stages began. Copies of all information letters and consent forms are attached as Appendix A.

3.4.5 Participants.

This research project involved four main participants and two class groups from one school site. The four participants were three secondary-level English/Literacy teachers and their principal, who also maintained an English/Literacy teaching workload. In the transcripts, these participants are designated with pseudonyms to ensure their anonymity. The teacher participant whose test administration was observed is presented as 'TT', as an acronym for 'testing teacher'. The other two teacher participants are presented as 'T1' and 'T2', as teachers one and two, the order of their interviews; the school principal is 'P'. TT and T2 are experienced teachers, having worked as

classroom teachers for 26 years when involved in the research. T1 was in his first year of teaching when involved in the research. T2 was the only female teacher participant. P was most experienced at the school level, who at the time of the data collection stage of the research was in his fourth decade of teaching. P was not involved in classroom test administration, but was involved as a school-level test administrator and taught students who were engaged in testing tasks.

The class groups observed during the standardised literacy test sessions were one Year Seven class and one Year Nine class, both co-educational class groups in a co-educational school. These class groups were ‘Humanities’ classes, the subject area that TT taught them. The Year Seven class group averaged between 18 and 23 students present during the three observed test sessions, and the Year Nine class group between 20 and 21 students present during the three observed test sessions. Both year groups generally had equal numbers of male and female students.

The three Year Nine test sessions were held during the final week of July, and the three Year Seven test sessions during the second week of August. TT had already taught the classes since the beginning of the school year in February of that year. Within the data, students’ names are also replaced with pseudonyms, indicating their year level, sex, student status, and a number to differentiate between students. For instance, ‘7FS1’ denotes the first Year Seven female student named in the Year Seven test sessions, and ‘9MS5’ denotes the fifth Year Nine male student named in the Year Nine test sessions.

3.4.6 Development of research instruments.

A combination of questionnaires, open-ended interviews and non-participant researcher observation was selected for the data collection instruments because of the predicted depth and scope of information they can afford. The use of multiple data collection instruments further allows flexibility for the researcher. Questionnaires, used first with teacher participants, begin the data collection by encouraging identification of initial ideas and points of interest, both by participants and the researcher. Such points of interest can then be further explored through even more flexible open-ended interviews, during which ideas can be discussed at more length and greater depth. Successive interviews allow ideas of interest to be discussed with participants to stimulate discussion. Non-participant researcher observation of the administration of standardised literacy test sessions, similarly, continues the trend of flexibility by allowing for the identification and generation of new ideas and theories, which can provide additional context for data interpretation.

3.4.7 Approval of research instruments.

During initial discussions and communication with the principal, the interview schedules and questionnaires (for teacher participants and the classes of students whose test sessions would be observed) were provided to the principal for approval. These interview and questionnaire schedules were also provided to teacher participants, with a note that the interviews would be conducted flexibly; avenues of interest that participants considered relevant could also be discussed during the interviews, and the order in which questions would be raised would be determined during the course of the interview session. At this

stage, research participants were again assured of their anonymity in the research, and that they would be provided copies of transcripts from sessions in which they participated for their approval before incorporation into the data analysis. Data instruments are attached as Appendix B.

3.5 Data Generation and Collection

Four forms of data were generated and collected to address the research questions. The data were not collected in chronological and exclusive order; for example, questionnaires and interviews were conducted before, during and after the observed test sessions. However, the basic order of the data collection was as follows:

- Testing session transcript data and the researcher's accompanying field notes
- Teacher and student questionnaire data
- Teacher and principal interview transcript data and the researcher's accompanying field notes.

The key forms of data analysed through the combined constructivist grounded theory and discourse analysis approaches were the data generated from transcripts of the audio-recorded standardised literacy testing sessions and transcripts of semi-structured participant (teacher and principal) interviews. These were selected as key data to be analysed due to the comparative richness and depth of detail contained within such data, and as a reflection of the study's focus on ways that test actors react within or as a result of the standardised literacy testing situation, including teachers' more extensive

considerations regarding their responses to this situation as a result of in-depth, open-ended discussions. The supplementary data (field notes and questionnaires) augment and enrich the testing session and interview data and analysis. Final copies of all data instruments are attached as appendices, in Appendix B.

The collected and analysed data were fieldwork data, which is common in social science research (Blaxter, Hughes, & Tight, 1996). These data were collected in the field, the school site as a natural social setting, and analysed in an attempt to “encounter life *as it happens* in the place or organisation where it usually occurs; to identify its patterns; and to produce an *understanding* of these” (Payne & Payne, 1994, p. 95; italics in original). This fieldwork-oriented approach is appropriate for research intended to be principally focused on classroom teaching perspectives and practices, as representative of school-based “natural setting/s” (Imel, Kerka, & Wonacott, 2002, p. 1).

The undertaking of fieldwork is also representative of the researcher’s theoretical positioning, an acknowledgement of the importance and value in allowing for a contextual interpretation of phenomena. Here, phenomena are considered as particular units of action; phenomena must be analysed “through direct interaction with, and interpretation by, the researcher” (Payne & Payne, p. 96). Such school research as this study aims to illuminate that “complex web that frames the everyday work of people in schools” (Brown, 2005, p. 33), without which, “we stand to forget just how complicated the teaching and learning process really is” (Brown, p. 33). The focus of this research is underpinned by an intent to explore and highlight the complexities and tensions

within standardised literacy testing, important given the popularity and commonality of this testing.

3.5.1 Testing session transcript data and field notes.

The testing session data (testing session transcripts, and field notes recorded concurrently with the audio recordings) were collected within the selected secondary school site. The data generation and collection stages involved the engagement of the researcher in non-participant observation in Years Seven and Nine standardised literacy testing sessions, of which there were three test sessions for each year/class group observed. Audio recordings were transcribed; TT was given the opportunity to review these verbatim transcripts and approve their inclusion into the data set.

The researcher recorded non-participant field notes during the standardised literacy testing sessions. During this observation, audio recordings and written notes were collected concurrently, to allow the researcher to capture more information and detail. These field notes are “the researcher’s detailed and descriptive record of the research experience, including observations, a reconstruction of dialogue, personal reflections, a physical description of the setting, and decisions made that alter or direct the research process” (Anderson, 2004, p. 128).

Throughout the observational note-taking, such detail came to include teacher and student behaviours and languages: the ways that TT introduced and administered standardised literacy tests to students; the routines TT followed during standardised literacy testing sessions, such as adhering to set

instructions, providing students with assistance, making context-specific decisions; and how the classroom actors involved in these testing activities behaved and interpreted the demands of the testing phenomenon. These primary points for observation represent the main patterns of TT and student action and interaction during the test sessions. The particular observations resulted from the researcher's decision to record what was taking place during observed test sessions. These observations were recorded in the field notes journal as a list of narrative-style observations, noting times of conversations, movements in the classroom and nuances of student reactions to the tasks.

Excerpts from the researcher's field notes will be provided throughout the analysis chapters, along with excerpts from the other collected data, to demonstrate the themes, patterns and issues identified in the constructivist grounded theory and discourse analysis stages. Excerpts from the researcher's field notes, as the other collected data, are also attached as Appendix C. These selected excerpts indicate the level and type of detail recorded during interviews and test sessions.

The test session observational data (both transcripts and field notes) were derived from naturalistic and non-participant observation, in the selected secondary-school classroom setting. The research literature advises that in order to limit the personal influence upon the observation data and upon the classroom setting, the researcher should endeavour not to manipulate or stimulate participant behaviour: the "behaviour is observed as the stream of actions and events as they naturally unfold" (Punch, 1998, p. 185), and the

researcher focuses on “[o]bserving a group in its natural operation rather than intervening in its activities” (Kumar, 2005, p. 121).

Observation is a common data collection technique in qualitative research, and unstructured non-participant observational data, used within this study, has been seen to offer a number of benefits and strengths. Unstructured non-participant observation allows the researcher/observer to witness and interpret that which is usually only experienced by those social actors who are natural participants in the phenomena. This is no guarantee that the researcher’s interpretations will reflect the interpretations of those being observed, nor is it a guarantee that the researcher will be able to avoid employing value-based language in interpretations or recorded notes (Nespor, 2006). In the case of recorded field notes, the purpose is to “make things *accessible*” (Nespor, p. 304; italics in original) and, in the case of transcribed test session data, the coding process attempts to make findings grounded and more real to participants’ presented experiences.

3.5.2 Questionnaire data.

During and immediately following the observation of the standardised literacy test sessions, questionnaires were distributed to the teacher participants. Student questionnaires were provided at the end of the final test session for each group. Questionnaires were used at these stages to transition into more focused participant data collection, and as a way of encouraging participant consideration of their opinions and uses of the tests prior to the interviews. The questionnaires were given to participants during the testing weeks to be

completed while participants' memories were fresher and while they were more likely to provide considered responses. Similarly, student questionnaires were given to students to complete immediately following each student group's final literacy test session. In addition to the benefits of immediate recollections, convenience was a factor: students were gathered as a group, equipped with writing materials at their desks and where the researcher could more easily collect completed questionnaires. The questionnaires are attached as Appendix B.

Questionnaires have a number of clear benefits as well as weaknesses. For example, respondents can be unwilling to provide detailed and revealing information, particularly if time-consuming; problems can arise from the distribution and return of the questionnaire instruments especially ensuring completed questionnaires; and responses can be lacking in depth. Gillham (2000) reminded that data from questionnaires "can appear (and usually are) thin, abstract and superficial" (p. 62). These issues affected this project, as participants chose to minimise responses and instead expand on ideas during interviews. The researcher needed to develop productive working relationships and remain in contact with participants to address and minimise problems with distribution and collection of questionnaires and to remind participants of potential details to be expanded later.

Although potentially problematic, questionnaires can be administered quickly and therefore be a relatively economical method of gathering preliminary data capable of informing later stages of data collection. They can be an effective way of identifying points of interest to contextualise or raise

during interviews. For example, in this research, participants were asked about the ways that they considered standardised literacy testing had positively affected their teaching, what aspects of standardised literacy testing they thought should be changed, and the role they perceived standardised literacy tests and testing play in contemporary education trends. These broad fields of questions served to inform later stages of this research.

Three different questionnaires were distributed in this study. Initial questionnaires were issued to the three teacher participants prior to the interviewing stage and during or immediately following the test sessions. In recognition of the need to cater for the levels of experience these teachers had with standardised literacy testing, two different teacher questionnaires were designed and provided. The wording was slightly different between these two questionnaires: for instance, instead of the questions probing teaching and testing experience, the questionnaire for the less experienced teacher probed perceptions and predictions. The questionnaires successfully garnered an initial overview of participant perspectives of standardised literacy testing and general literacy assessment, and encouraged their critical responses about what they do, and why, with standardised literacy testing. Examination of the questionnaire responses is incorporated in Chapter Five.

The third questionnaire used was given to the two observed class groups for completion immediately following their three standardised literacy testing sessions: reading/comprehension, writing/composition and spelling. Fourteen questionnaires were collected from the observed Year Seven class group, and 17 from the Year Nine class group. The student questionnaire

sought information about their perspectives on standardised literacy tests and their opinions on the ways in which these tests are used by teachers, and to what effect. Student questionnaire responses effectively communicated student perceptions of standardised literacy testing and literacy assessment, and gave voice to student perspectives of literacy assessment practices.

The data generated from these student questionnaires was chiefly used to supplement the data from the observational journal, participant interviews, and testing session transcript data. The student data, while of relevance, are not the study's main focus; the main focus is teachers' perceptions and practices regarding standardised literacy testing. The student questionnaires fulfilled research objectives of gathering potentially useful data that could highlight the possible affect on students and the classroom environment of standardised literacy testing. In addition, this source of data was sought as comparison and contrast with teacher data for a fuller appreciation of the different 'realities' within the classroom testing site.

Analysis of the student and teacher questionnaires, as already stated, is incorporated in Chapters Five and Six. The questionnaires help to enrich the data analysis by contributing to the contextualisation, development and analysis of the themes identified within the data, as required for examination focused on the research questions.

3.5.3 Interview data.

As a data collection tool, interviews are strongly represented in qualitative educational research, and are commonly utilised by researchers for the

collection of rich data to develop grounded understanding. Of relevance to this research, semi-structured interviews can be an important data generation and collection tool, and “can be the richest single source of data” (Gillham, 2000, p. 65) because of the “ ‘richness’ of the communication that is possible” (Gillham, p. 62). Such interviews allow the researcher to access actors’ “experiences and subjective views” (Whipp, 1998, p. 54). Interviews allow participants to “reveal the personal framework of their beliefs and the rationales that inform their actions” (Whipp, p. 54) through simultaneously providing “[d]etailed, vivid, and inclusive accounts of events and processes” (Whipp, p. 54). In this research, the interviews were conducted as “interviews in which the intent is to understand informants on their own terms and how they make meanings of their own lives, experiences, and cognitive processes” (Brenner, 2006, p. 357). The researcher recognised the very flexibility of the interview to perceive of or identify “new dimensions of a problem or to discover clues that connect its different elements” (Whipp, p. 54). In this study, detail illuminated through the observed test sessions was further discussed with all teacher participants. This allowed participants to suggest different interpretations and dimensions, validated other participants’ experiences, and grounded analysis in the data.

While semi-structured interviews can provide flexibility for the interviewee and allow deeper data to be generated, planning and preparation was necessary to achieve full benefits of the method. The researcher in this study planned and prepared the interviews prior to the data collection and arrived at the school site with tentative topics, deliberative questions and

progressive sequence for the interviews. This structure was balanced with a degree of flexibility, and the data collection allowed for deviation from the initial structure. This balance of structure and flexibility was employed in an attempt to generate richer interview data responsive to research participants while maintaining a general focus on the topic.

Through the interviews, the research participants, secondary-school teachers of Years Seven and Nine English/Literacy, were asked both open- and closed- questions, with the majority of questions being open-ended. Responses to interview questions recorded by audiotape and transcribed verbatim provided the researcher with “in-depth information” (Kumar, 2005, p. 134). The researcher was aware, however, that data generated from open-ended questions can present a challenge. Open-ended questions can be more difficult than closed-question responses for comparison and analysis purposes. This was partially overcome through the intensive, line-by-line coding, which enabled the complex and at times abstract themes to emerge. The detailed information attained through the use of open-ended questions and subsequent careful coding therefore considerably strengthened the analysis. The potential difficulty of comparing the data was also addressed by engaging in the initial and focused analysis stages according to the data type: interview data were analysed in isolation from the testing session data, to avoid confusing or blurring the emerging issues in the two data types. The interview and test session categories were combined and compared in the theoretical coding stage.

Semi-structured interviews enabled the voices of interviewees and research participants to be valued and incorporated in this research. This research was encouraged in a dialogic or praxis direction, with more personal and detailed information being obtained than would have been possible with structured and inflexible interviews. The researcher recognised that semi-structured interviewing enables respondents' active involvement in the "construction and validation of meaning" (Lather, 1991, p. 63), ensuring that the "critical inquiry" (Lather, p. 63) be a "fundamentally dialogic and mutually-educative enterprise" (Lather, p. 63). Research involvement will not automatically result in beneficial outcomes for participants though, as has been explained in the literature, "emancipatory intent is no guarantee of an emancipatory outcome" (Acker, Barry, & Essevold, in Lather, p. 80). It is acknowledged that in the context of this research, which is reliant upon volunteers for the interviews, participants might have pre-existing and strong opinions about standardised literacy testing and literacy assessment. It is possible that these participants could desire an outlet for their personal and professional opinions, and an active part in the research.

Contributing to the construction and validation of meaning through participation in research could be part of a conscious activist agenda using research as instrumental to change. This background could be viewed as both potential limitation and strength. It is argued that genuine participant involvement in the research can ensure the real grounding of the data, findings, recommendations and conclusions in the participants' realities of the testing phenomena. The nexus between lived experience of the classroom and testing

could be more fully explored. This research therefore supports greater responsiveness than that produced by a de-contextualised approach.

Initial P and TT interviews took place before the test sessions, to gather background on the student cohort and school context. The main teacher participant interviews were conducted after the collection of the testing session transcripts and questionnaires. This was because the emerging patterns, themes and issues generated through the initial questionnaires and test sessions were subsequently raised and discussed with the participants in the interviews. Collection and analysis of the multiple forms of data prior to the interview stage, importantly, facilitated a contextual reference between the researcher and interview participants that would allow for more higher-level data to be generated through the interviews.

The semi-structured interviews were audio recorded, and transcripts typed verbatim; these were submitted to each research participant involved for correction, removal and agreement. The transcripts were initially analysed through a constructivist grounded theory coding analysis approach, then a discourse analysis approach at the later analysis stages. The interview programme is attached as Appendix B.

3.6 Approaches to Data Analysis

Within the employed case study design, this study draws on combined analytic techniques from constructivist grounded theory (Charmaz, 2000, 2001, 2005, 2006; Charmaz & Mitchell, 2002) and discourse analysis (Carabine, 2001; Fairclough, 2001; Taylor, 2001a, 2001b) approaches. As is the case within a

constructivist grounded theory approach, the data collection and analysis processes are deliberately entwined to ensure the generation of theory from the actual data. This is consistent with the description of case study research by Anderson (2004), who stated “data collection and data analysis are concurrent activities” (p. 154). Also consistent with a broad case study approach, the data sources and techniques in this research are broad, to allow for the emergence and recognition of issues of interest (Anderson, 2004; Charmaz, 2005, 2006; Charmaz & Mitchell; Dey, 1999, 2007; Haig, 1995; Whipp, 1998; Yin, 2006).

Combining relevant elements of analysis approaches and perspectives can enable better awareness of the data, the topic of research and best method for analysis (Phillips and Jørgensen, 2002). Constructivist grounded theory and discourse analysis can be effectively employed together, due to their flexibility of use and intent (Carabine, 2001; Charmaz, 2000, 2001, 2005, 2006; Charmaz & Mitchell, 2002; Taylor, 2001a, 2001b). Research that utilises constructivist grounded theory and discourse analyses approaches can enable the generation of theory and findings that clarify and examine participant and researcher perspectives, question underlying assumptions, and highlight and challenge hegemonic practices, through a careful and theoretically-aware application of the tenets and steps of the particular analysis approaches (Carabine; Charmaz; Charmaz & Mitchell; Taylor). Research utilising these approaches enables the generation of theory that is grounded in the data, meanings and conclusions are transparent to the participant and reader, and relevant for field practitioners and researchers alike (Charmaz; Taylor).

The adoption of more than one methodological approach in the data analysis stages makes the links clearer between the themes, issues and discourses in both the collected data and in the literature. This approach enables a more thorough critique of the data and the meanings contained therein. These methods allow the data to be examined in contrasting but complementary ways. The combination of the constructivist grounded theory and discourse analysis approaches provides clear processes for coding, patterning, and theoretically exploring the themes and ‘big D’ discourses, and for examining the effects, consequences and manifestations of these dominant discourses.

The analysis processes employed in this research depend upon and utilise the contextual awareness developed through the literature review. The researcher was aware that the study must be situated “within the body of related literature” (Stern, 2007, p. 123) in order to give proper “credit to other researchers [and] demonstrate how you built upon it so that you can see further” (Stern, p. 123). Such credit is accorded to other researchers and their work by clarifying these contributions to the field and expanding on the extant knowledge, incorporated in Chapters Six and Seven.

3.6.1 Constructivist grounded theory.

A constructivist grounded theory approach to research is an inductive method for theory generation, drawing theory from the data rather than the data serving to verify a preconceived hypothesis or theory, working from the theory generation rather than the theory verification perspective (Charmaz, 2006;

Punch, 1998; Saldaña, 2009). “Theory emerges from ... and is said to be ‘grounded’ in” the data” (Whipp, 1998, p. 52). The data collection and analysis occur simultaneously, as a means of refining and shaping the emergent theories and ensuring these theories are grounded in the data (Charmaz & Mitchell, 2002; Corbin & Strauss, 1990). Within this study, a simplified constructivist grounded theory coding analysis is employed, in which four main coding stages are used, to construct the codes, categories, themes and discourses. These four main stages are termed initial, focused, axial and theoretical coding, respectively (Charmaz, 2006; Saldaña).

During the initial coding stage, the data are examined in close detail (line-by-line), to construct the first stage of codes. These codes are kept active, to emphasise the focus on what is happening in the data (Charmaz, 2005, 2006). Using active codes “curbs our tendencies to make conceptual leaps and to adopt extant theories *before* we have done the necessary analytic work” (Charmaz, 2006, p. 49; emphasis in original). These active codes are the “beginning” stage of the coding analysis that enables the researcher to “fracture or split the data into individually coded segments” (Saldaña, 2009, p. 42). These individually coded segments represent the first stage of the coding, from which patterns and meanings in the data can eventually be identified.

The first three codes constructed from each of the two codes data types in this research are:

<p>Interview transcript codes 1-3:</p> <p><i>Identifies differences in teachers' methods of test administration (code 1)</i></p> <p><i>Presents tests as isolated activities, no preparation (code 2)</i></p> <p><i>Identifies tests used as diagnostic tools to identify student weaknesses (code 3)</i></p>
<p>Testing session transcript codes 1-3:</p> <p><i>Presents tests in a positive way (code 1)</i></p> <p><i>Presents tests in a negative way (code 2)</i></p> <p><i>Presents tests as useful/helpful for students and learning (code 3)</i></p>

Figure 3-1. *Codes 1-3, interview transcripts and testing session transcripts.*

Codes serve a deeper purpose than merely being tools or terms used by the researcher for differentiating distinct or related meanings, or being helpful labels given to units of meaning identified as important in the data (Dey, 1999). These initial, active codes begin the process of categorising and conceptualising the patterns of action and meaning. These codes are read and re-read through constant comparison against the data and the other codes, to ensure they conceptualise what is happening in the data. This in turn assists in grounding future coding stages in the data. Initial coding is “open-ended” that allows the “researcher to reflect deeply on the contents and nuances of your data and to begin taking ownership of them” (Saldaña, 2009, p. 81).

The initial codes are the result of the researcher’s examination of the participant interview and testing session transcript data. These data were examined for what Charmaz (2006) refers to as “sensitizing concepts” (p. 17). These sensitising concepts in this particular coding analysis reflect the research questions: standardised literacy testing value and validity; participants’ perceptions and interpretations of standardised literacy testing; approaches to test administration; and patterns of classroom interaction as related to testing.

It is not only these concepts that are identified in the coding stages, however, as any emergent points of interest are followed, then later refined and revised.

Once the data are coded, the initial codes are concentrated and grouped together into categories, to show the processes and patterns identified in the initial codes. In this research, the stage of sorting codes into categories is termed focused coding. Focused coding is more analytical and theoretical than the initial coding because it requires the analyst to reflexively interact with the data and the active codes. The categories constructed from the focused coding are “more directed, selective, and conceptual” than those constructed from the initial coding phase, and enable the synthesis and examination of “larger segments of data” (Charmaz, 2006, p. 57). This is more theoretical and conceptual than the initial coding as this stage necessitates critical thinking about what the initial codes mean and how they relate.

The first categories and the first three codes that constitute them from each of the two codes data types are:

Classroom control (interview category 1)	<i>Refers to SLT administration rules (code 14)</i>
	<i>Stresses/identifies/refers to the role/importance of classroom control in lesson and testing sessions (code 15)</i>
	<i>Refers to poor student behaviour – influential upon teaching and administration (code 17)</i>
Pragmatics of standardised literacy testing administration, and administrative uses (test session category 1)	<i>Over-identifies administration as feature/issue of tests (code 16)</i>
	<i>Refers to the administrative uses of test data (code 17)</i>
	<i>Reads aloud test instructions (code 36)</i>

Figure 3-2. *Category 1 (interview and test session), with three codes.*

In this study, as already stated, frequency counts and calculations of codes in the data were undertaken in this and the following stages, and comparison of themes and behaviours was also included. It is at this stage that the coding begins to highlight the “processes” in the data (Charmaz, 2006, p. 51).

Processes and categorisation are refined and revised in the third coding stage, which further conceptualises the data. The third step of the constructivist grounded theory coding analysis employed in this research is here termed axial coding, so called because the coding levels and data are re-examined and conceptualised around the core, central ‘axes’ (Saldaña, 2009). This coding stage deconstructs the seven participant interview categories and nine testing session categories, and effectively “describes a category’s properties and dimensions and explores how the categories and subcategories relate to each other” (Saldaña, p. 151).

The categories are broken up and examined, to make the data “denser, more complex and more precise” (Charmaz & Mitchell, 2002, p. 515). This stage allows for the identification of the themes from the data and the codes, in order to conceptualise and represent the “conceptual elements of [the] theory” (Glaser & Strauss, 1967, p. 36, cited in Dey, 1999, p. 48). At this stage, the coding from both main data types is synthesised, so the themes represent the axial coding of both data types already patterned into codes and categories. The first of these themes and the categories that constitutes it is:

<i>Maintenance of control, behaviour management, order</i> (theme 1)	Classroom control (interview category 1)
	Teacher as authority (test session category 9)

Figure 3-3. *Theme 1 with categories.*

Following the construction of the themes, the properties and parameters of the themes are illuminated and clarified in Chapter Five through the process of examining how the themes address the research questions. It is from this stage of the analysis that the dominant discourses are drawn. The theoretical coding stage breaks down the themes to clarify and question the broader patterns of meaning-making, identity, power and knowledge regarding standardised literacy test administration for these participants and at this school site are clarified. At this stage of the coding, the analysis “progresses toward discovering the central/core category that identifies the primary theme of the research” (Saldaña, 2009, p. 151), which in this research means the dominant discourses. This adapted theoretical coding stage is undertaken through the process of examining the themes through the lens of the research questions.

‘Theoretical saturation’ is the term used for the situation in which “the data are sufficient to make and justify an interesting argument” (Phillips & Hardy, 2002, p. 74). In this study, theoretical saturation was reached when the identified codes and themes had ‘saturated’ the identifiable in the data by the researcher; the codes and themes were sufficiently thorough, detailed and identifiable (through quotes and examples) so that further attempts to code and compare the data highlighted no new themes regarding the topic. This process

of iterative coding involves returning to data and codes to ensure the thoroughness, saturation and grounding of coding. This process aided the identification of theoretical patterns within and across the levels of coding, and helped to highlight key themes and dominant discourses, particularly through the later coding stages during which the codes were refined, compared, revised and broken down to reveal more complex patterns within the data.

As an iterative coding process, the data and levels of codes are read and re-read to deepen understanding and appreciation of the content, patterns and themes within the data. This iterative nature of constructivist grounded theory coding qualifies it as a constant comparative approach (Charmaz, 2005, 2006). The comparative and iterative nature of this analysis, which allows reinterpretation and rediscovery, together with a poststructuralist-influenced questioning of data and of the researcher's self (reflexivity), helps to "make fundamental processes explicit, render hidden assumptions visible, and give participants new insights" (Charmaz, 2006, p. 55).

The constructivist grounded theory coding stages allow researchers to identify themes and issues throughout collected data, locate points of comparison and group together theoretically similar codes and categories. In this way, the coding meant that in this research the data were thoroughly examined for themes and issues, participants' interpretations, arguments and justifications, and test-related behaviours, processes and types of interactions.

The researcher employing this analysis approach uses memos, recorded alongside the data and coding, which can be utilised to determine and analyse the deeper context and meaning of the data, codes, categories and thematic

areas. This research adopted the use of memo records during the data generation, collection and analysis stages. Memos allow the researcher to record interesting theories and ideas about each segment of data or the data as a whole, observations that would otherwise be forgotten or lost. Memos also serve as a reminder of these initial theories and ideas and of the context from which they were taken, so redrawing the researcher into the data and the participants' realities.

3.6.2 Discourse analysis.

As discourse analysis is not a single method of analysis but rather a broad field of research and analysis, there is not one way of conducting a discourse analysis, and the researcher must select the most appropriate approach for the research field, philosophies, interest and objectives (Naples, 2003; Phillips & Hardy, 2002). Discourse analysis “contains, first, philosophical (ontological and epistemological) premises regarding the role of language in the social construction of the world, second, theoretical models, third, methodological guidelines for how to approach a research domain, and fourth, specific techniques for analysis” (Phillips & Jørgensen, 2002, p. 4). As C. Lewis (2006) noted, however, such a discourse analysis is not an end in itself, and the discursive analyst still needs research questions and purposes.

The discourse analysis was not conducted in order to ‘find’ assumed discourses or practices; the discourse analysis steps/stages were not presumed to highlight “inevitably oppressive” discourses (Naples, 2003, p. 145). The lenses lent by Carabine’s (2001) and Fairclough’s (2001) approaches enabled

the researcher to examine for discourses and themes, not necessarily find them; “familiarity with such a range of theory and empirical work gives you tools to see more and say more about the particular things you’re studying” (Nespor, 2006, p. 299). This perspective was in keeping with the employment of a constructivist grounded theory coding analysis, which is inductive and takes theory from the data rather than deductively seek to verify or refute pre-conceived theory through the data.

Discourse analysis, as employed in this study, is concerned with the broader issues and themes identified through the language (test session and interview transcripts) that represent the more overarching patterns of power, meaning-making, interpretation, coping and expectation, for instance. Again, in this way, the discourse analysis and its findings build on the results of the coding analysis, by considering the theoretical issues and discursive patterns in an ideologically-influenced frame of analysis.

This adopted discourse analysis approach questions the ways that the practices reinforce or challenge hegemonies and norms, the social implications of the discourses within the wider power contexts, representations and sites of contestation (Carabine; Fairclough; Phillips & Hardy, 2002; Rowan, 2001; Taylor, 2001a, 2001b). This discourse analysis approach also questions and problematises forms, manifestations and effects of power, important because “power reaches into the very grain of individuals ... inserts itself into their actions and attitudes, their discourses, learning processes and everyday lives” (Foucault, 1980, p. 39).

Within this study, the content of collected data are examined, following the constructivist grounded theory coding approach, to construct and highlight “broad social, cultural, and ideological processes” (Bloome & Clark, 2006, p. 227), in this case the larger phenomena and tensions suggested through the participants’ accounts of standardised literacy testing. The understanding of, and approach to, discourses and discourse analysis underpinning this study provides the tools to look beyond the linguistic meanings and, instead, highlight, interpret and examine the social practices and tensions within the participants’ articulated experiences of the testing phenomenon.

The school site where the data were collected allows exploration of an evident “subject of inquiry” and clarifies an “obvious discursive struggle so that discursive activity was clearly evident and likely to be linked to ways in which individual actors sought to protect their interests” (Phillips & Hardy, 2002, p. 67). Some discourses will be more dominant, powerful, influential and authoritarian than others (Carabine, 2001; Fairclough, 2001). The school site shows the ways in which test-related discourses are “productive in that they have power outcomes or effects. They define and establish what is ‘truth’ at particular moments” (Carabine, p. 268).

Carabine’s (2001) discourse analysis steps adapted and employed in this research are:

- 1 Select your topic – identify possible sources of data
- 2 Know your data
- 3 Identify codes, categories, themes and dominant discourses
- 4 Look for evidence of an inter-relationship between discourses
- 5 Identify the discursive strategies and techniques that are employed
- 6 Look for absences and silences
- 7 Look for resistances and counter-discourses
- 8 Identify the effects of the discourse
- 9 Context 1 – Outline the background to the issue
- 10 Context 2 – Contextualise the material in the power/knowledge networks
- 11 Be aware of the limitations of the research, your data and sources. (Carabine, 2001, p. 281)

Figure 3-4. *Carabine's (2001) adapted discourse analysis steps.*

The first three steps were undertaken through the coding analysis, and within step 4, interrelationship between key themes as well as discourses increased critical, initial understanding of the data concepts and patterns, facilitating the coding stages. The remaining discourse analysis steps were not necessarily examined and addressed in the order listed above, as the particular elements or aspects of the dominant discourses were not necessarily be identifiable in any set order.

Across the discourse analysis was necessitated an identification and contextualisation of the ways that the discourses are perpetuated at the school

level of standardised literacy testing and shape participants' experiences and realities. It also involved questioning what is not present, necessitating the researcher's actions "to follow up gaps in analysis: to identify what is inferred but not mentioned; expected but absent; to identify limitations of dominant discourses" (Budd, 2009, pp. 93-94). Absences and silences can be as revealing as what is explicitly present, as they can indicate hidden or taken-for-granted assumptions and shed further light on the effects of discourses, as well as highlight reasons for the silencing of voices or ideas (Taras, 2007).

Identification of examples in the data of counter-discourses and of resistances to dominant discourses was important as these provide interesting opposition to normalised practices, expectations and opinions. Here, participants' attempts to challenge or subvert the dominant discourses were examined, as were their effect on the dominant discourses or contribution to identified discursive tensions and struggle. The effects of the discourses identify and examine the ways that they shape participants' and possibly other test actors' experiences of the testing discourses. In examining this aspect, broader contexts were considered also, as the discourses effects on the contextual framework.

Context was considered in chapters one and two. Chapter One outlined the background of standardised literacy testing in Tasmanian secondary schools, and Chapter Two expanded on this by positioning the research in the context of debates regarding standardised literacy testing, the politics surrounding and permeating the area in its broad sense. Chapters six and seven, through analysis and findings, increasingly expand and examine the

contextual issues pertaining to the research. Research limitations are discussed in Chapter Seven. This includes critical discussion of the limitations of the study, including possible limitations resulting from the researcher's influence. This stage also raises possible methodological, theoretical and practical considerations for other research projects.

In addition to Carabine's (2001) work, Fairclough's (2001) discourse analysis steps were influential. This research adapted and employed these stages:

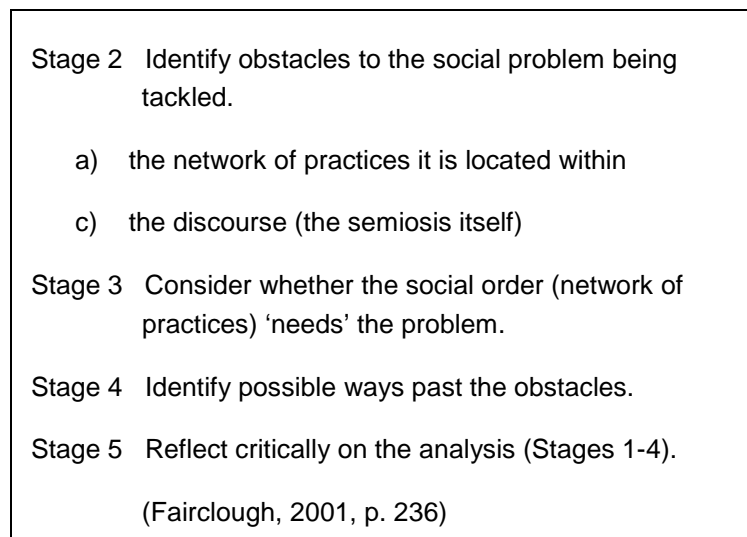


Figure 3-5. *Fairclough's (2001) adapted discourse analysis steps.*

In the context of this research, the social problem is an educational phenomenon that represents an educational, perhaps social and political, problematic for test actors. Further, from its extension in politics and society in Tasmania, the phenomenon has potential for wide social and political effects. The 'problem' of issues within teacher administration of standardised literacy testing was examined in terms of the broader networks of practices

within which standardised literacy testing is located in an educational sense. Alongside the identification of the ‘problem’ is the questioning of who benefits from the current state, and why. This means examining the purposes served by those issues participants recognised within standardised literacy testing, and by participants’ chosen ways of reacting and responding to the standardised literacy testing phenomenon.

This discourse analysis approach “looks for hitherto unrealized possibilities for change in the way social life is currently organized” (Fairclough, 2001, p. 236), to identify ways that the data and the broader literature and results of other research presented solutions or alternatives. This is explored in the final two chapters, particularly in the study recommendations (Chapter Seven). The recommendations and research limitations address the final stages of this discourse analysis approach, as also the above-mentioned step 11 of the Carabine discourse analysis approach.

3.7 Structure of Analysis Chapters

The details and results of the initial, focused and axial coding analysis stages are presented in Chapter Four, the details and results of the adapted theoretical coding stage are presented in Chapter Five through the structure of the research questions, and the findings of the discourse analysis are presented in Chapter Six. As noted in Chapter One, use of the term ‘findings’ (employed in relation to the discourse analysis) is a deliberate choice intended to highlight the increasing level of critical and contextual analysis in that stage of the study as compared with that of the constructivist grounded theory analysis.

In Chapter Five: Research Questions and Themes, data excerpts are closely examined to highlight the patterns and themes and illustrate the results drawn. The research questions are explored through the themes and parameters that constitute them, and what the data say about the research questions is reiterated. In that chapter, the results of the constructivist grounded theory coding analysis are structured through the research questions.

The findings of the discourse analysis – the discursive patterning and struggles, and their causes, manifestations and repercussions – are presented in Chapter Six. That chapter presents the findings of the discourse analysis, overviewing, examining and comparing the dominant discourses and discursive tensions along power/knowledge networks and in light of the broader literature corpus.

Throughout all chapters, as noted in Chapter One, emphasis in data extracts and examples represents emphasis from the original. This means that emphasis identified in the transcript, by the speaker (in the case of audio recorded transcripts) or by the author (in questionnaire responses and field notes).

Chapter Four: Coding Analysis of Data

Introduction

This chapter presents the results of the first three coding stages of the constructivist grounded theory analysis of the key data (Charmaz, 2006; Saldaña, 2009). These data were collected and generated at a secondary-level State school during the preparation for, administration of and follow up from the mandated standardised literacy testing. The data were collected and generated in order to examine the study's dual focus: on teachers' professional opinions of standardised literacy testing, and on processes and practices that can be considered to take place during teachers' administration of the tests. This dual focus provides the basis for the two research questions, 'How do teachers perceive and interpret the value and validity of standardised literacy testing?' and 'What patterns of action and interaction characterise standardised literacy testing?' These two research questions are returned to in Chapter Five, in which the results of the final stage of the coding analysis are presented.

The previous chapter, Chapter Three: Methodology, described the analysis approaches employed in the research, and theoretical principles that underpinned the selection and adoption of the analysis approaches. As outlined in the previous chapter, four types of data were collected: transcripts from interviews with participants (teachers responsible for standardised literacy test administration, and their principal); transcripts from testing sessions; the researcher's field notes; and questionnaires from participants (teachers) and the two class groups of students whose testing was observed.

Examples from the interview transcripts and testing session transcript transcripts are presented in this chapter, and examples from all four data types are presented as relevant in the following chapters. Extracts from transcripts and data are also attached as Appendix C.

Only the interview and test session transcript data were analysed along the constructivist grounded theory coding processes, as these were the most ‘open’ data types collected and the key forms of data used in this analysis. The field notes and questionnaire responses were used in identifying the discourses, expanding upon the initial results of the coding analysis. The inclusion of these additional data sources, including the researcher’s observational notes, enables the analysis to expand upon identified patterns and be informed by comparisons between participants’ voices.

This chapter is in five sections. The first section, 4.1 Details of Coding Analysis, briefly overviews the three constructivist grounded theory analysis coding stages employed through the chapter. Sections 4.2 to 4.4 present the three coding stages, the third stage (4.4) synthesising the analysis of the two key data types. These sections list and explain the codes, categories and themes constructed from the three coding stages (initial, focused and axial, respectively) of the interview and testing session transcripts to demonstrate the results and contextualise the later findings drawn from the data and contextualised in broader literature.

4.1 Details of Coding Analysis

The results of the coding analysis presented in this chapter overview the initial, focused and axial coding stages of the transcripts from six audio-taped and transcribed participant interviews (three with TT, and one each with T1, T2 and P), and six observed and audio-recorded standardised literacy testing sessions (three from the Year Seven class group, and three from the Year Nine class group).

The interview and test session transcript data were analysed and successively grouped into codes, categories and key themes, following the general steps for initial, focused and axial coding of the constructivist grounded theory coding approach adopted within this research (Charmaz, 2001, 2005, 2006). Initial coding is the term for the stage of grouping data into codes, focused coding the term for the grouping of codes into categories, and axial coding the re-examination of codes and categories and their re-patterning into key themes. An example is provided in the following Figure 4-1, with an example of an extract from the data showing the way that a sentence or grouping of sentences was designated with a code descriptor numbered for ease of coding and comparisons, and sorted into a category and a theme.

<p>“We’re – we would be intrigued to know what sort of mathematical process they use for this equation, and it would be a fun statistical process that they use. I know the people who do it, (...) and (...), they’re number crunchers and they know what they’re doing.”</p> <p>P, Interview 1, p. 5, L 10-13</p>	
<i>Memo</i>	<i>Code, Category and Theme</i>
The Principal’s uncertainty as to the ways that the results are computed doesn’t necessarily mean that he doubts the accuracy of the result itself, but rather that he is unaware of the processes undergone in order to statistically measure and represent students’ literacy levels.	<i>Expresses a lack of understanding as to the formulation of data from tests or ways test results are computed (code 24)</i>
Instances of this code can be patterned within broader critiques of standardised literacy testing and tests as this code identifies an issue within testing that can be addressed. In the case of this code, the issue can be addressed by increasing school-site test actors’ understanding and awareness of what happens between the time that students take their tests and when the statistical label of literacy level is returned to the school. It is not a critique or flaw within the classroom site of test administration, but part of a larger-scale process of standardised literacy testing.	Broader test critiques (category 5)
This code and category, once broken down and examined for meaning and effect, pertains most particularly with the practical side of standardised literacy testing. The code’s emphasis is uncertainty of one element of the testing processes, and the category’s emphasis is on a problematic but solvable element of testing at a macro-administrative level. The code and category can be refocused to a central theme of what happens during broad testing processes, including an uncertainty regarding what happens or in what ways it happens.	<i>Pragmatics of standardised literacy tests</i> (theme 3)

Figure 4-1. *Coding example.*

Figure 4-1, above, shows the level of conceptualisation and patterning of the data required to construct and refine the levels of coding. The

importance of memoing during the coding process is also indicated through the above. Memos were used during the coding of the data to record initial ideas regarding meanings of data, note and compare patterns across coded data, and prompt for clues in 'reading' and coding the data. Memos allowed the researcher/analyst to record ideas specific to coding stages and data, which in turn proved helpful during the back-and-forth processes of coding.

Coding is an iterative and not a linear process, meaning in this analysis that new codes were constantly identified and added, and newer and later codes were also applied to the data (Charmaz, 2005, 2006; Saldaña, 2009). The coding analysis frequently revealed previously unidentified themes that were often relevant to data that had already been analysed and coded at initial stages. Further, while some codes were clearly related and very similar in topic or focus as a result of the relatively small number of codes, each of the codes referred to specific aspects or themes identified within the data.

The final coding labels were determined only after the exhaustion of the coding processes and the point of theoretical saturation. Altogether, the coding process constructed 44 separate initial codes and seven categories from the participant interview transcripts, and 39 separate initial codes and nine categories from the testing session transcripts. From these 16 combined categories, five themes were constructed. These codes, categories and themes are presented in table format in the coming sections, and are overviewed in more detail in this chapter.

As already stated, the researcher's field notes and questionnaires were not coded in the same format as the participant interview data and testing

session transcript data, although extracts from them are presented in the following chapter and attached as Appendix C. Field notes and questionnaire data were used to supplement the analysis of interview and testing session data, to highlight absences and silences, resistance and counterdiscourses, and agreement and similarity.

4.2 Initial Coding

Initial, line-by-line coding is the first step in this constructivist grounded theory analysis (Charmaz, 2005, 2006; Dey, 2007). As already explained, the data analysis must occur simultaneously within the data collection within a constructivist grounded theory study; initial coding in this research, therefore, was undertaken at the very beginning of the dual collection and analysis process. As this stage was iterative, the constructed initial codes are the result of frequent revision and refinement (Saldaña, 2009).

Within the initial coding stage, codes were both named and numbered, aiding presentation and differentiation. During the initial coding stage, the codes were noted by page and document, for comparison across individual data documents for emergent points of interest. While undertaking this stage, for instance, every instance within the interview transcripts that the researcher recognised that TT identified differences in teachers' methods of test administration, the margin beside the passage was marked with '1'. This code was similarly noted beside all other instances to enable faster identification of initial code patterns. This tallying of codes was used specifically as a tool for the location and highlighting of themes, and to compare frequencies of codes.

4.2.1 Initial coding: Participant interviews.

The coded participant interviews, as previously explained, were represented by the transcripts from the six interviews conducted with the participants (TT, T1, T2 and P). The data from the initial participant interview transcript analysis were categorised into 44 different initial codes, which highlighted participant opinions regarding standardised literacy testing and their roles within the testing processes and practices within their school site. These constructed codes and their numerical frequency across the six participant interview transcripts are listed below in Table 4-1.

Table 4-1

Interview transcripts: Codes and frequencies.

<i>Code no.</i>	<i>Code Descriptor</i>	<i>Frequency</i>
1	Identifies differences in teachers' methods of test administration	32
2	Presents tests as isolated activities, no preparation	14
3	Identifies tests used as diagnostic tools to identify student weaknesses	42
4	Identifies tests used as a check for students' internal school results	32
5	Suggests differences in teachers' use of test data	9
6	Presents self as unconcerned about testing or the tests' administration	6
7	Refers to broader school-wide uses of test data rather than as diagnostic tools for individual students	15
8	Refers to students within context of school catchment – judging students according to perceived lower abilities, behaviours and work ethic	33
9	Refers to student skills, abilities, learning levels/standards	80
10	Refers to difficulties in addressing literacy problems that are identified in test data within a student group with mixed ability levels	15
11	Presents that the reality of standardised literacy testing is different to the perceived ideal	71
12	Presents tests as valuable but not prepared for	3
13	Suggests test success is valuable to/for students	10
14	Refers to standardised literacy testing administration rules	5
15	Stresses/identifies/refers to the role/importance of classroom control in lesson and testing sessions	36
16	Suggests extraneous community problems as potentially influential upon students	6
17	Refers to poor student behaviour – influential upon teaching and administration	36
18	Refers to non-testing teaching and learning activities	35
19	Refers to methods/approaches of teaching styles and classroom lesson administration	31
20	Suggests testing and assessments are not necessarily the best indicator of student intellect or effort	29
21	Refers to specific tasks or learning skills (e.g., punctuation, spelling)	11

22	Refers to helping students with specific tasks or problems	13
23	Presents standardised literacy tests as administered (by schools) in a confusing or disorganised way	34
24	Expresses a lack of understanding as to the formulation of data from tests or ways test results are computed	8
25	Refers to tests or testing data in a positive way	24
26	Explicitly identifies language and wording as a particular flaw in test documents or structure	9
27	Identifies problems in standardised literacy testing processes – reporting, releasing of data to teachers, etc.	34
28	Identifies student skill/ability as influential on their attitudes to/performance in testing sessions	42
29	Suggests the idea the tests and the test data are used for political purposes	20
30	Identifies difficulty in providing help so as not to answer questions for students	14
31	Refers to the validity of test responses or data	31
32	Suggests teachers consider tests to be of limited value	18
33	Identifies ways the tests could be improved	14
34	Claims/suggests teachers administer tests properly and responsibly	19
35	Portrays the tests as difficult for students (educationally)	32
36	Claims/suggests teachers are uninfluential on test data and responses	2
37	Identifies teacher influences on test data and responses	5
38	Suggests a variety of influential factors on student literacy standards or performance in standardised literacy testing	19
39	Identifies teacher performance as relevant element of school - specifically	4
40	Identifies differences in gender behaviour of teachers	4
41	Reference to classroom or school (administration) rules – requirements on students and teachers	12
42	Identifies and refers to student difficulties in undertaking specific tasks (not just testing)	19
43	Justifies not following test administration rules	11
44	Presents test administration and testing as unpleasant or stressful for teachers	8
<i>Total</i>		<i>947</i>

947 instances of the 44 initial codes, as per above Figure 4-2, were identified across the six participant interview transcripts. These codes highlighted the different ways that participants described and considered various aspects of standardised literacy testing. These initial codes constructed from the participant interviews informed the research about participant opinions of standardised literacy testing and of what happens during test administration. These served as ‘sensitising concepts’ (Charmaz, 2006, pp. 16-17) and highlighted further issues and points of interest, such as references to participants’ professional interpretation of students’ experiences of testing and of expectations and requirements upon test actors (both teachers, as test administrators, and students, as test-takers).

The comparative frequencies of codes across the interviews indicate the test elements and issues that participants most often discussed. Table 4-2 below lists the codes noted in the top ten most frequent instances across the interview transcripts.

Table 4-2

Interview transcripts: Most frequent codes.

<i>Code no.</i>	<i>Code Descriptor</i>	<i>Frequency</i>
9	Refers to student skills, abilities, learning levels/standards	80
11	Presents that the reality of standardised literacy testing is different to the perceived ideal	71
3	Identifies tests used as diagnostic tools to identify student weaknesses	42
28	Identifies student skill/ability as influential on their attitudes to/performance in testing sessions	42
15	Stresses/identifies/refers to the role/importance of classroom control in lesson and testing sessions	36
17	Refers to poor student behaviour – influential upon teaching and administration	36
18	Refers to non-testing teaching and learning activities	35
23	Presents standardised literacy tests as administered (by schools) in a confusing or disorganised way	34
27	Identifies problems in standardised literacy testing processes – reporting, releasing of data to teachers, etc.	34
8	Refers to students within context of school catchment – judging students according to perceived lower abilities, behaviours and work ethic	33

The most noted code, ‘Refers to student skills, abilities, learning levels/standards’ (code 9), refers to teachers’ references to both behavioural and academic abilities. This common code suggests that teachers considered student abilities an important factor in testing and test administration, as does the third/fourth most noted code, ‘Identifies student skill/ability as influential on their attitudes to/performance in testing sessions’ (code 28). Indeed, the frequency of ‘Stresses/identifies/refers to the role/importance of classroom control in lesson and testing sessions’ (code 15) and ‘Refers to poor student behaviour – influential upon teaching and administration’ (code 17) continue this trend across the interview coding.

The comparative frequencies also highlight what the participants claimed to be the uses of test data: 'Identifies test used as diagnostic tools to identify student weaknesses' (code 3), 'Identifies tests used as a check for students' internal school results' (code 4), 'Refers to broader school-wide uses of test data rather than as diagnostic tools for individual students' (code 7), and 'Suggests differences in teachers' use of test data' (code 5).

Table 4-2 above shows the spread of code frequencies and particularly noticeable is the gap between the frequencies of codes 9 and 11 (first and second most noted codes) and codes 3 and 28 (both the third/fourth most noted). Code 9 is noted 80 times across the interviews, code 11 noted 71 times, and codes 3 and 28 are both noted 42 times. This coding tally indicates that participants overall were preoccupied with considerations of student and classroom needs and demands.

A number of these initial codes from participant interviews presented participants' observations of how testing took place at their school site. Codes 14, 15, 17, 23, 34 and 41 pointed to early similarities in participants' consideration and articulation of the element of test administration, despite some differences. Participants' focus on administration rules included openly discussing with the researcher the particular rules of administering test sessions (code 14), the expected participant recognition of the need for control during test sessions (code 15) and the overarching understanding of requirements upon themselves and their students to comply with rules (code 41). By their expressed identification of the disruption caused by student misbehaviour during tests (code 17), the participants communicated the context and

problematised for the purposes of this research. Code 34 indicated that participants could consider their administration of the tests to be proper and responsible, yet code 23 presented the view that school-wide test administration and organisation was not as well done as possible. Considered together, these codes refer not just to rules and expectations dictated by the testing documents, but also to the school rules that govern how teachers and students are to behave and react to situations (protocols to follow, for instance).

Codes 36 and 37 highlighted a noteworthy pattern, that of participants' perceptions of their own, rather than their students', influences upon student test responses and test data. Code 36 showed instances in which participants claimed they had no influence over testing; code 37 shows instances in which participants claimed otherwise. Code 30 showed the frequency with which participants articulated one area of struggle for them during test sessions, that of being unable to help students who found tests difficult.

Codes 1 and 5 highlighted participants' acknowledgement of differences of teacher opinions regarding the ways that teachers engaged with testing: code 1 showed the participant opinion that teachers administered the tests in different ways (different processes followed and guidelines used, for instance), while code 5 showed the participants' opinion that teachers made differences out of test data. Some of these different uses made by the participants of the test data can be seen through initial coding to be: diagnostic, to identify areas of student literacy skill that need further help (code 3); and to validate teachers' assessment judgements, as a cross-check against students' internal grades (code 4). In the initial coding, participants at times also

presented test data as used more for broader school-wide purposes, such as planning whole year level literacy programmes or computing change in literacy levels over time (code 7). Participant opinion also raised with concern test data being used for political purposes, such as school funding or holding schools accountable, rather than for providing teachers and schools with literacy information (code 29).

4.2.2 Initial coding: Testing session transcripts.

The coded testing session transcripts, as previously explained, are the transcripts of the six standardised literacy testing sessions observed and recorded at the school site. The data from the initial testing session transcript analysis were categorised into 39 different initial codes, which highlighted patterns of action and interaction at the classroom site of standardised literacy testing. Table 4-3, as follows, presents these codes and their numerical frequency across the six testing session transcripts.

Table 4-3

Testing session transcripts: Codes and frequencies.

<i>Code no.</i>	<i>Code Descriptor</i>	<i>Frequency</i>
1	Presents tests in a positive way	3
2	Presents tests in a negative way	54
3	Presents tests as useful/helpful for students and learning	2
4	Presents tests as useful in the administration of learning topics for teachers	4
5	Presents tests as isolated incidents or tools	23
6	Presents tests as not thoroughly utilised in student learning or the administration of learning topics	9
7	Presents tests as problematically administered	44
8	Presents tests as potentially useful in student learning and the administration of learning topics	13
9	Presents tests as positive incidents/tools but in a negative way	5
10	Presents self as noncommittal, blasé or uncaring as to the tests, their uses and/or their administration	26
11	Displays uncertainty as to the uses and administration of the tests	17
12	Presents tests as beneficial for employment and post-school	4
13	Presents tests as beneficial, specifically to student learning outcomes	10
14	Presents tests as challenging and difficult	99
15	Presents tests as an indicator of learning progress	22
16	Over-identifies administration as feature/issue of tests	87
17	Refers to the administrative uses of test data	3
18	Refers to students as 'client-base', 'clientele'	1
19	Constructs student ability/interest as influential upon the administration or running of tests	48
20	Presents or refers to the practice of tests as different to the perceived ideal	21
21	Identifies classroom control as major testing element	139
22	Categorises/judges schools on behavioural issues based on student catchment	6
23	Refers to student (poor) behaviour as influential on delivery	42
24	Presents tests as revealing of student difficulties	15

25	Refers to/discusses non-/post-testing activities (daily report, report cards, post-testing silent reading, questionnaires)	9
26	Refers to student skills	39
27	Presents self in position of power in test administration	7
28	Presents self in position of impotence in test administration	25
29	Controls student behaviour	267
30	Explains researcher presence in room	9
31	Encourages student interest in tasks	35
32	Explains specific aspects of skills	4
33	Uses humour	31
34	Responds positively to student rudeness	12
35	Allows student to act as test-refusor, belligerent	25
36	Reads aloud test instructions	164
37	Offers help on tasks	142
38	Dismisses student test queries	52
39	Encourages student	17
<i>Total</i>		1535

The codes, as listed in Table 4-3 above, were teacher-oriented in order to capture and reflect the focus of this study on the ways that action and interaction took place while TT was administering the test sessions, and on TT's expressed or implied standardised literacy testing experiences and perceptions regarding standardised literacy testing. There were 1535 distinct instances in total relating to the 39 codes. Highlights at this initial stage included TT's opinions on the tests, the ways the tests were administered in the classrooms and the data used, and the particular merits and flaws in the standardised literacy test design, implementation and application. A number of codes referred specifically to behaviours particular to test administration, TT's responses to student approaches to undertaking the tests, instructions given for

the students in taking the tests, and teacher-student interaction during test administration.

The spread of frequencies contained much interesting information for this stage of the research. The following Table, Table 4-4, lists the codes noted in the top ten most frequent instances across the interview transcripts.

Table 4-4

Testing session transcripts: Most frequent codes.

<i>Code no.</i>	<i>Code Descriptor</i>	<i>Frequency</i>
29	Controls student behaviour	267
36	Reads aloud test instructions	164
37	Offers help on tasks	142
21	Identifies classroom control as major testing element	139
14	Presents tests as challenging and difficult	99
16	Over-identifies administration as feature/issue of tests	87
2	Presents tests in a negative way	54
38	Dismisses student test queries	52
19	Constructs student ability/interest as influential upon the administration or running of tests	48
7	Presents tests as problematically administered	44

As highlighted in the above Table, ‘Controls student behaviour’ (code 29) was notable, particularly because of its frequency: it had by far the most common occurrence within the testing session transcript data. ‘Reads aloud test instructions’ (code 36) and ‘Offers help on tasks’ (code 37), too, indicate the importance placed during the observed test sessions of ensuring a smooth flow of test activities, the former code showing TT’s necessary compliance with the expectation upon him as administrator to explain tasks, the latter

showing his attempts to ensure that students understood and could complete the given test tasks. Similarly, 'Identifies classroom control as major testing element' (code 21) was notable for its frequency within the data transcripts, as fourth most frequent code, a code that too highlighted the importance for TT of controlling factors within the classroom that could disrupt or affect the test session. This is evidently important as the fifth most noted code, 'Presents tests as challenging and difficult' (code 14), noted 99 times across the six transcripts, suggests TT's opinion about his students' capacity to complete tasks without careful teacher oversight. Also noteworthy were the comparative frequencies of 'Encourages student' (code 39), at 17 instances, and 'Controls student behaviour' (code 29), the most noted code at 267 instances.

Each initial testing session transcript code, while related to others, was specific in its particular meaning. Codes 3, 4, 6 and 8 provided examples where different meanings could be indicated from seemingly minor wording differences. Codes 3 and 4 present TT's opinion that standardised literacy tests were helpful or useful, but in two distinctly different areas of teaching and learning: for student learning progress (code 3), and for teachers' administration of learning topics (code 4). Both codes 6 and 8, in contrast, referred to these two areas (student learning and teacher administration of learning topics), yet differed in what standardised literacy tests meant for those ends: potentially useful for these ends (code 8), or not thoroughly utilised for these ends (code 6). Codes 12 and 13 again retain their different foci through precise wording: while code 13 referred specifically to academic benefits of testing, such as improvement of student literacy skills, code 12 referred more

to broader, non-academic benefits from particular aspects of standardised literacy testing, such as personal gain or improved employment options for students.

4.2.3 Summary of codes.

The initial coding stage was the first stage of the constructivist grounded theory coding analysis (Charmaz, 2001, 2005, 2006). During this stage, the six interview transcripts and six testing session transcripts were analysed through an iterative, line-by-line coding process, to highlight the active codes within the transcripts. This initial coding stage resulted in the construction of 44 initial codes from the interview transcripts and 39 initial codes from the testing session transcripts.

The initial codes from across these two key data types highlight the actions and statements recorded in the interview and testing session transcripts in relation to standardised literacy testing, the tests' administration, roles of and expectations upon test actors, opinions of the tests, and the like. These initial codes provided initial insight into what the data say about standardised literacy testing at this school site. These initial codes were then examined along the second coding stage, that of focused coding, by which they were grouped into patterns.

4.3 Focused coding

Within this analysis, the focused coding stage is a process of “developing categories without distracted attention ... to their properties and dimensions” (Saldaña, 2009, p. 155). In so doing, the categories constructed from the

interview and testing session transcript data “explicate ideas, events, or processes in your data” and “may subsume common themes and patterns in several codes” (Charmaz, 2006, p. 91). This process also necessitates some decision-making, in reviewing already coded and grouped material with fresher insight. This section lists the categories constructed through the focused coding stage and explains the patterns identified through the focused coding stage. This section presents in depth five categories selected for their particular importance in further coding stages: each of these five categories was grouped into each of the five themes, and together they present a picture of the focused coding stage of both the interview transcripts and the testing session transcripts.

4.3.1 Focused coding: Participant interviews.

The 44 initial participant interview transcript codes were grouped into patterns of similarity and meaning, from which seven categories were constructed.

Figure 4-2, as follows, tabulates these participant interview transcript categories and the codes that constructed them. The seven categories represented the first level of interpretation and deeper reading of the initial line-by-line codes, and showed the variety of participant concerns regarding a variety of issues, both testing- and teaching-related. Following the Figure, two categories from the participant interviews, with their codes and patterning, are detailed, to provide an example of the patterning and processes involved in this stage of the coding of the interview transcripts.





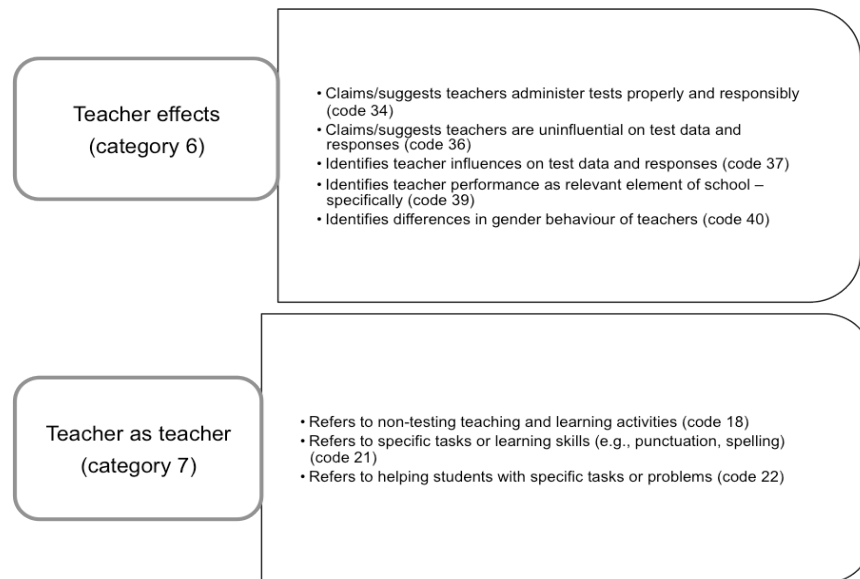


Figure 4-2. *Interview transcripts: Codes and categories.*

The seven categories constructed from the focused coding of the participant interview transcripts showed the broad groupings of participants' opinions and accounts of testing at their school site. Participants expressed various criticisms and critiques of standardised literacy tests and testing, regarding test documents, processes, requirements, relevance, appropriateness, uses, usefulness, value and validity. Participants also discussed issues not specific to standardised literacy testing, such as common teaching expectations and behaviours, presented opinions of their effects upon their students' progress (during testing and in a broader sense), and communicated judgements of their students according to student catchment and student academic and behavioural abilities.

4.3.1.1 Interview category 3: Judgements of students.

Category 3, 'Judgements of students', was constructed from 'Refers to students within context of school catchment – judging students according to perceived lower abilities, behaviours and work ethic' (code 8), 'Refers to student skills, abilities, learning levels/standards' (code 9), 'Suggests extraneous community problems as potentially influential upon students' (code 16), 'Identifies student skill/ability as influential on their attitudes to/performance in testing sessions' (code 28), 'Portrays the tests as difficult for students (educationally)' (code 35), 'Suggests a variety of influential factors on student literacy standards or performance in standardised literacy testing' (code 38), and 'Identifies and refers to student difficulties in undertaking specific tasks (not just testing)' (code 42).

These seven codes are patterned through the participants' comments about students' capacity to succeed in the test situation. One grouping within this broad category pattern highlights participants' interpretations of student academic and behavioural abilities, as affecting student ease in undertaking test and non-test tasks. Examples of this category pattern include: *'To a certain extent, it doesn't matter how you present test-like activities, to a particular group of kids, they're not going to want to do it'* (TT, Int. 3, p. 2); *'they were such a difficult class that um it was almost impossible to say, well, 'yes, these results actually mean something', because they really weren't concentrating, and some of them were giving one another a hard time, and some of them weren't listening when we did the spelling bit'* (T2, Int. 1, p. 2); and *'7FSI,*

interestingly – I reckon she didn't put any time into her score, because 7FS1's not dumb. Into her performance' (P, Int. 1, p. 3).

The second grouping within this category pattern is that of extra-school effects: the indicated judgement of students' abilities according to school catchment, and community influences upon students performance. Examples of this category pattern include: *'we're dealing with a client-base here where work ethic isn't high, in a lot of the kids'* (TT, Int. 1, p. 1); *'there are issues to do with extraneous problems that might be happening in the community, which can affect the reactions of individuals on any particular day'* (TT, Int. 1, p. 2); *'a lot of that depends on the background of the individuals you are dealing with, in that if I was working in another school, with a different group of clients'* (TT, Int. 2, p. 3); and *'comparing us with [State all-girls secondary school] is like comparing a cheese stick and a carrot, I mean, they're not the same thing at all'* (T2, Int. 1, p. 12).

Within this broad category grouping is evident participants' implied judgement that the school catchment was in a number of ways less academically and behaviourally capable as the student groups at other schools in the general area. This further suggested that students' test results would reflect the students' comparative lower abilities, and as such, administering the tests to these students necessitated awareness of the effects of student behaviour and ability on test sessions and results. The participants' references to the academic difficulty of test tasks for this student group, and to the student group's 'skills, abilities, learning levels/standards' (code 9), indicate the importance of this pattern.

4.3.1.2 Interview category 4: Issues in test administration.

Category 4, 'Issues in test administration', was constructed from 'Identifies differences in teachers' methods of test administration' (code 1), 'Presents tests as isolated activities, no preparation' (code 2), 'Presents self as unconcerned about testing or the tests' administration' (code 6), 'Presents tests as valuable but not prepared for' (code 12), 'Presents standardised literacy tests as administered (by schools) in a confusing or disorganised way' (code 23), 'Identifies difficulty in providing help so as not to answer questions for students' (code 30), and 'Justifies not following test administration rules' (code 43).

These seven codes were grouped together through their shared pattern of similarity in the participants' identification of often-problematic elements in test administration, with varying implied causes and consequences. The first of these pertains to participants' presentation of the test planning and preparation stage at the school site. This pattern communicates the participants' consideration that, regardless of the value or importance of the testing, the tests are not sufficiently planned or prepared for before they reach the classroom administration stage. Examples of this category pattern include: *'no grade nines Monday, it won't be Monday. It will have to be Tuesday, I suspect Tuesday first thing is the best time to do one, Wednesday first thing will have to be another, and then there's Thursday. It's possible, by the way the timetable is set up, that we might actually have to split ...'* (TT, Int. 2, p. 3); and *'you do*

have to apologise, because it's a lot to do in a one-week space, with the numeracy as well' (T2, Int. 1, p. 2).

The second category pattern pertains to participants' expressions of their classroom test administration. Within this second pattern, participants confirmed that teachers administer the tests differently, and are not altogether concerned with this component of the testing. Examples of this category pattern include: *'every teacher does it differently. So that what I do might be completely different to what the person in the room next door is doing'* (TT, Int. 1, p. 1); *'you have to choose how you administer the test'* (TT, Int. 2, p. 2); *'I read through them first, then I read through them to the kids. And then we picked out the bits, and then we – I did those bits again'* (T1, Int. 1, p. 9); and *'if you just did what it says, the kids would have been quite confused, I think. So I don't follow every single thing to the letter'* (T2, Int. 1, p. 1).

Within this second category pattern, participants articulate the difficulty they experience in balancing providing help to students within their role as teacher and test administrator without invalidating students' test responses, and go so far as to provide reasons and justifications for administering tests in a way that they identify is noncompliant with expectations. Examples of this category pattern include: *'So, it's just necessary, it's expedient, to provide more than the basic instructions that are provided in the test itself. And if you didn't do that, then you would just be setting yourself up for disaster. In doing that, you try very hard not to influence the result that the individual gets'* (TT, Int. 3, p. 5); *'you gotta give everybody the opportunity, and if that means reinterpreting the wording of the question, then I don't see that as*

unreasonable' (TT, Int. 3, p. 6); and '*Trying to make sure that they can do the best they can on the test, yes*' (T2, Int. 1, p. 3).

4.3.2 Focused coding: Testing session transcripts.

As per the process with focused coding of participant interviews, the 39 initial testing session transcript codes were grouped into patterns resulting in the construction of nine categories. Figure 4-3 as follows tabulates these testing session transcript categories and the codes that constructed them, and following the table, three test session categories, with their codes and patterns, are detailed, to provide an example of the patterning and processes involved in this stage of the coding of the testing session transcripts.

Pragmatics of standardised literacy testing administration, and administrative uses (category 1)	<ul style="list-style-type: none"> • Over-identifies administration as feature/issue of tests (code 16) • Refers to the administrative uses of test data (code 17) • Reads aloud test instructions (code 36)
Teacher indecision as to the benefits of standardised literacy testing (category 2)	<ul style="list-style-type: none"> • Presents tests as potentially useful in student learning and the administration of learning topics (code 8) • Presents tests as positive incidents/tools but in a negative way (code 9) • Presents self as noncommittal, blasé or uncaring as to the tests, their uses and/or their administration (code 10) • Displays uncertainty as to the uses and administration of the tests (code 11)
Positive correlation between standardised literacy testing and student learning outcomes (category 3)	<ul style="list-style-type: none"> • Presents tests in a positive way (code 1) • Presents tests as useful/helpful for students and learning (code 3) • Presents tests as useful in the administration of learning topics for teachers (code 4) • Presents tests as beneficial, specifically to student learning outcomes (code 13) • Presents tests as an indicator of learning progress (code 15)
Standardised literacy testing as negative, problematic, lacking in educational merit (category 4)	<ul style="list-style-type: none"> • Presents tests in a negative way (code 2) • Presents tests as isolated incidents or tools (code 5) • Presents tests as not thoroughly utilised in student learning or the administration of learning topics (code 6) • Presents tests as problematically administered (code 7) • Presents or refers to the practice of tests as different to the perceived ideal (code 20)
Judgements on student communities and student catchment (category 5)	<ul style="list-style-type: none"> • Presents tests as beneficial for employment and post-school (code 12) • Refers to students as 'client-base', 'clientele' (code 18) • Categorises/judges schools on behavioural issues based on student catchment (code 22)
Judgements on student standards, learning levels (category 6)	<ul style="list-style-type: none"> • Presents tests as challenging and difficult (code 14) • Presents tests as revealing of student difficulties (code 24) • Refers to student skills (code 26)

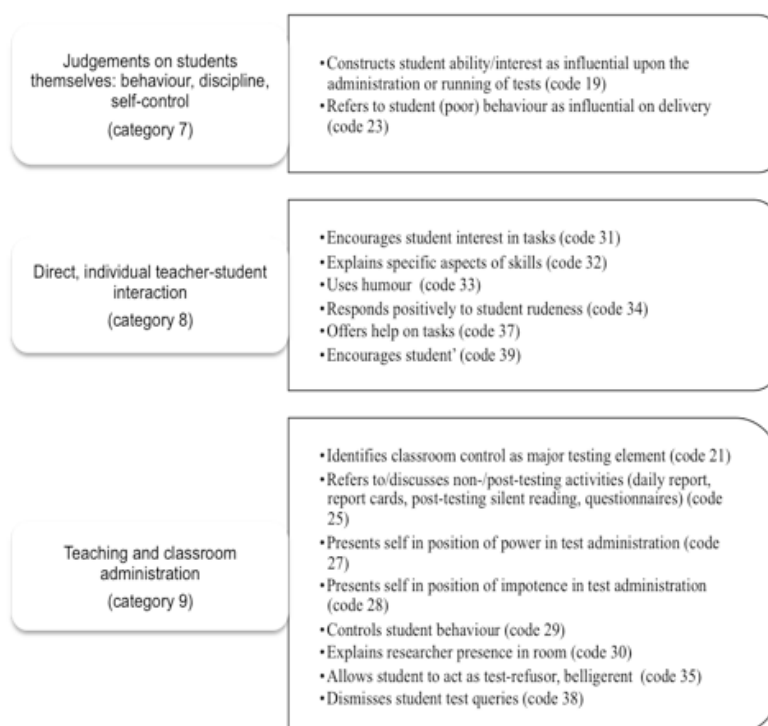


Figure 4-3. *Testing session transcripts: Codes and categories.*

The nine testing session categories represented a level of grouping, interpretation and deeper reading of the initial codes. These categories showed the classroom manifestations of participant-identified issues, related to both testing and to broader teaching issues. The categories that resulted from this focused coding of the test session transcripts highlighted the administrative challenges and demands induced by the tests, differing opinions of the value, validity and benefits of standardised literacy tests and testing, various judgements made or expressed about the student group, instances and specifics of interaction during test sessions, and examples of TT's teaching and classroom management style.

4.3.2.1 *Testing category 2: Teacher indecision on the benefits of standardised literacy testing.*

Category 2, ‘Teacher indecision as to the benefits of standardised literacy testing’, is constructed from ‘Presents tests as potentially useful in student learning and the administration of learning topics’ (code 8), ‘Presents tests as positive incidents/tools but in a negative way’ (code 9), ‘Presents self as noncommittal, blasé or uncaring as to the tests, their uses and/or their administration’ (code 10), and ‘Displays uncertainty as to the uses and administration of the tests’ (code 11). These four codes are again grouped because of the pattern of TT indecision concerning perceived benefits and problems of standardised literacy testing, specifically test administration and test data. The codes within this category indicated some of the problems in the identification or perceptions of standardised literacy testing as either completely positive or negative.

These codes showed that TT appeared to interpret many aspects of standardised literacy testing activities as being either both positive and negative, or as being too difficult to decide between the two. TT’s lack of knowledge and concern about the uses and administration of standardised literacy tests were examples of this general pattern of indecision; this uncertainty as to testing dates, procedures and practices, and his evident lack of concern as to this uncertainty, also suggested an uncertainty about the tests’ benefits. Examples of this category pattern are: ‘*What do the instructions say – just have a look under the bit that says ‘Comprehension’’* (Year Nine Reading Test, p. 2); ‘*When do you do this one? / Um, probably on – tomorrow morning, first thing*’ (Year Nine Reading Test, p. 5); ‘*Oops, we don’t need answer sheets*

this morning' (Year Nine Spelling Test, p. 1); *'Right, pencils – actually, I've got to check to see if we need pencil or pen'* (Year Nine Spelling Test, p. 2); and *'Use a pen if you want, it doesn't matter. Long as it's readable'* (Year Nine Spelling Test, p. 3).

The evident and frequent displays of TT confusion as to whether these tests are positive or negative demonstrated a pattern of ambivalence or uncertainty; for example, he occasionally used negative language and arguments while seemingly intending to appear positive about the tests. His perspective of standardised literacy tests and standardised literacy testing, therefore, appeared confused and ambiguous. Similarly, he made veiled remarks about testing; for example, he seemed positive, but communicated negativity in the way he talked and as a result confused students as to why they should do the tests. Some TT comments about the tests seemed to be positive references about the benefits of standardised literacy testing, however both the context and vocal tone and expression used by him could have been identified as cleverly worded criticisms of the ways the tests do and do not actually help students. Examples of this category pattern are: *'the easiest and the quickest of the tests'* (Year Nine Spelling Test, p. 1); *'This, you will be pleased to know, is the last of the tests'* (Year Nine Spelling Test, p. 2).

4.3.2.2 *Testing category 8: Direct, individual teacher-student interaction.*

Category 8, 'Direct, individual teacher-student interaction', was constructed from 'Encourages student interest in tasks' (code 31), 'Explains specific aspects of skills' (code 32), 'Uses humour' (code 33), 'Responds positively to

student rudeness' (code 34), 'Offers help on tasks' (code 37), and 'Encourages student' (code 39). This category, which drew on six codes, presented patterns from the data of teacher-student interaction and behaviours and provides numerous examples. Modes of communication-, behaviour- and response-based interaction as specific examples within this pattern include the reiteration of activity schedules during testing session times, TT explanation of activities and skills, teacher response to attempts by students at negotiation over activities and tasks, and procedures followed with individual students during classroom time.

TT's encouragement of students was a strong pattern within this category of direct teacher-student interaction. Teacher attempts at encouraging student interest in testing tasks by explaining that the students could successfully complete the test was one such interaction, and TT used explicit explanation of specific aspects of skills needed as part of the encouragement. Individualised encouragement of students was shared by coded data in this pattern. TT sometimes offered help to students struggling with tests during testing sessions, including both solicited and unsolicited assistance by the teacher to students. TT also gave some students encouragement both during and after testing sessions concerning their perceived performance and behaviour. Humour was employed during test sessions when interacting with students as interactions with individual students, and with students at a group level. Examples of this pattern are: *'all you have to do is work out how to spell 'em. Or do your best at spelling them'* (Year Seven Spelling Test, p. 3); *'There's no rush 'cos we've got plenty of time, we've got next lesson to finish'*

(Year Nine Writing Test [session 1 of 2], p. 5); *'How are you going? You're doing a good job. Got plenty of time. Have a break if you need to'* (Year Nine Reading Test, p. 7); and *'Who-ho-ho, that was good! Eight pencils! Throwing pencils at you!'* (Year Nine Spelling Test, p. 3).

These codes showed a further pattern of TT's responses to student misbehaviour and rudeness, suggesting his attempts to foster positive teacher-student relationships and avert more disruptive classroom disturbance. This pattern showed instances in which TT elected to focus on a positive teacher-student relationship and immediate classroom order rather than his authority being asserted in an excessively aggressive manner. Within this pattern was evident TT effectively choosing to relationship-build or appease students using non-confrontational (positive) methods rather than impose (negative) discipline were common in his direct, individual interaction with students. Examples of this pattern are: *'Some of the words have been left out. / Der'* (Year Seven Spelling Test, p. 4); *'Bullshit. I'm not very happy about that'* (Year Nine Writing Test [Session 1 of 2], p. 4); and *'Don't worry about that. / Give us a look'* (Year Nine Reading Test, p. 4).

4.3.2.3 *Testing category 9: Teaching and classroom administration.*

Category 9, 'Teaching and classroom administration' was also constructed from a number of codes, these being 'Identifies classroom control as major testing element' (code 21), 'Refers to/discusses non-/post-testing activities (daily report, report cards, post-testing silent reading, questionnaires)' (code 25), 'Presents self in position of power in test administration' (code 27), 'Presents self in position of impotence in test administration' (code 28),

‘Controls student behaviour’ (code 29), ‘Explains researcher presence in room’ (code 30), ‘Allows student to act as test-refusor, belligerent’ (code 35), and ‘Dismisses student test queries’ (code 38). These codes revealed a pattern that showed TT’s direct teaching actions and references to classroom administration.

TT’s perceptions of his classroom role as a teacher and administrator of standardised literacy tests were reflected in each of these eight codes. Similarly, the codes shared examples relating to his responsibilities with regards students, teaching-learning-assessment and standardised literacy testing tasks. Such perceptions, as suggested through the codes, affected and shaped his pedagogical and theoretical understandings, thereby strongly influencing his methods of classroom teaching and standardised literacy testing.

Within this category pattern, TT provided direction for students finishing the test, selecting and directing the allowed activity, students’ behaviour, when it would occur and for which students. Examples of this pattern include: *‘No you’re not, you’re going to sit and you’re gonna do what you’re told’* (Year Seven Reading Test, p. 8); and *‘I’ll do a little deal with you, if you do your best with this, [AP] said you can go to her and do some computer stuff’* (Year Seven Writing Test, p. 2).

The pattern of TT’s assertion of teaching and testing authority was also evidenced by the teacher’s verbal behaviour management and student discipline, demonstrated frequently in the testing sessions. Examples of this pattern include: *‘You do not speak, you do not spend any time worrying about what other people are doing, you are worried only about what you are doing’*

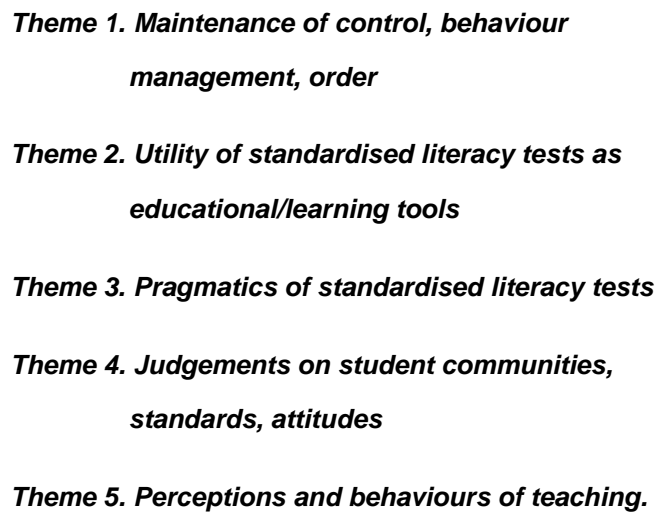
(Year Nine Writing Test, p. 1); *'Right well you be quiet and you listen'* (Year Nine Writing Test, p. 3); *'No, just need to wait, please. No boys I asked you to shift the desks and I meant it. Do not move the furniture. 9MS4, move it back, please! Don't mind who moved it, I'm asking you to move it back, please'* (Year Nine Spelling Test, p. 1); *'7MS4 – 7MS4, go and sit yourself over on the decisions desk please, over there'* (Year Seven Writing Test, p. 2); and *'That means no conversations from now on, thanks. Not one sound'* (Year Seven Spelling Test, p. 4).

4.3.3 Summary of categories.

In this section, the focused coding stage of the constructivist grounded theory coding analysis has been clarified. The five categories explored in depth serve as examples of the processes engaged in during this coding stage, from which the themes were then identified. The categories highlight patterns across participants' explored experiences and opinions of standardised literacy testing and observed test sessions. The five categories explored are examples of the themes, as each of these five were sorted into each of the five themes.

4.4 Axial coding

The research process during the axial coding stage involved the breaking down of both key data types. The 16 emergent categories were compared against each other for the overarching or core themes and again checked against the data. This further iterative process generated the themes from the data and the results of the initial and focused coding stages. From this exhaustive process, the axial coding stage resulted in five themes:



***Theme 1. Maintenance of control, behaviour
management, order***

***Theme 2. Utility of standardised literacy tests as
educational/learning tools***

Theme 3. Pragmatics of standardised literacy tests

***Theme 4. Judgements on student communities,
standards, attitudes***

Theme 5. Perceptions and behaviours of teaching.

Figure 4-4. *Themes.*

These themes reflected the broader recurrent issues within the data, and showed overlap between the participant interviews and the testing session transcripts in terms of the overarching patterns. Figure 4-5 below shows the categories that constituted these themes, and amalgamates the results of the initial and focused coding of both key data types to present and engage with a synthesised, data-inclusive and grounded analysis. From this stage in the analysis, the key data types will be analysed together, because the coding levels effectively brought together the key data types enabling a deep examination of amalgamated issues, patterns, themes and concepts. These themes are critically examined in Chapter Five, so the detail provided in this chapter about the themes is an indicator of the broad patterns, parameters and concepts that constitute them.

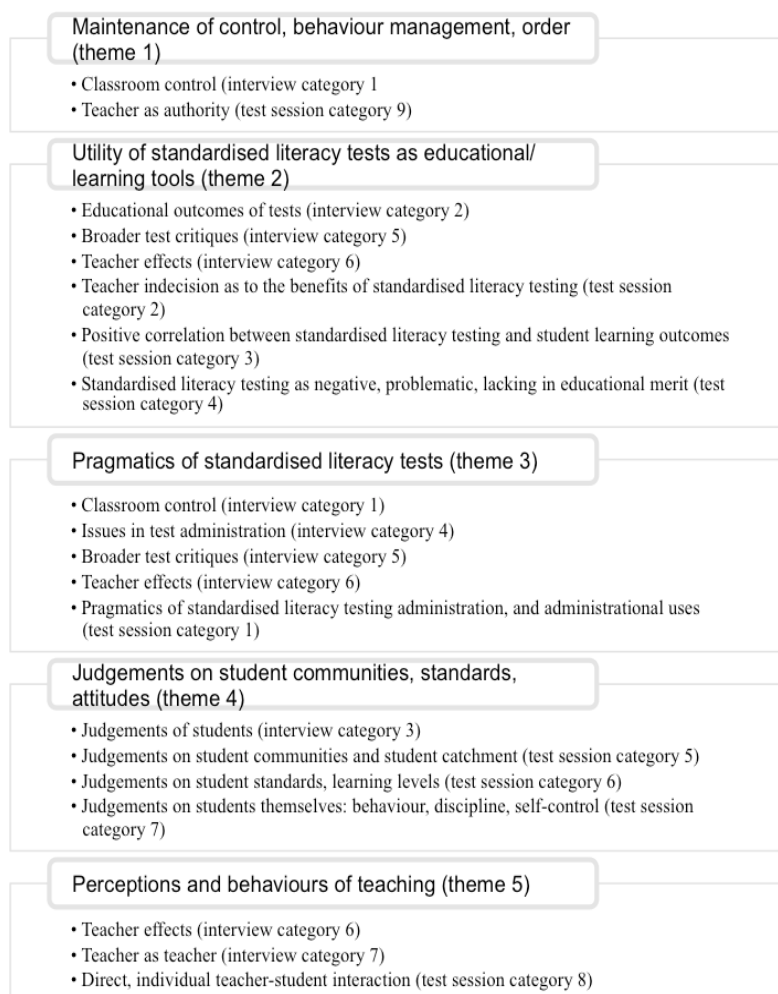


Figure 4-5. *Themes and categories.*

4.4.1 Theme 1: Maintenance of control, behaviour management, order.

The first theme, ‘Maintenance of control, behaviour management, order’, consists of only two categories, ‘Classroom control’ (interview category 1) and ‘Teacher as authority’ (test session category 9). The central theme connecting these codes and categories was the importance that participant teachers accorded classroom management. Within this theme, the data communicated that effective classroom management was highly important, in order to complete testing tasks and administer learning activities.

As a major pattern within this theme, the participants considered that their student group frequently demonstrated disruptive and violent behaviours. Such behaviours necessitated a particular teaching approach for the classroom and school environment to be ones in which learning tasks could be completed and teachers and students could work within a conflict-free situation. The central theme here was broader than simply a pattern of participant references and opinions regarding the need for a particular approach to student discipline and behaviour management. Rather, a pattern is reflected of participant interpretations of their students' behavioural needs and of how the participants reacted to or addressed them. The data clarified participants' perceptions and TT's actions regarding what is the most effective way of interacting with students and reasons why, interpreting and responding to students' behavioural demands or expectations, and considerations of the ways that they as teachers affected their classroom sites.

4.4.2 Theme 2: Utility of standardised literacy tests as educational/learning tools.

The second theme, 'Utility of standardised literacy tests as educational/learning tools', was constructed from six categories. These were: 'Educational outcomes of tests' (interview category 2), 'Broader test critiques' (interview category 5), 'Teacher effects' (interview category 6), 'Teacher indecision as to the benefits of standardised literacy testing' (test session category 2), 'Positive correlation between standardised literacy testing and student learning outcomes' (test session category 3), and 'Standardised literacy testing as negative, problematic, lacking in educational merit' (test session category 4).

These six categories all relate around the central theme of standardised literacy testing and tests as an educational and learning tool, though the categories reflected the different properties within this theme.

This central theme of the utility of standardised literacy tests as educational/learning tools incorporates participants' opinions as well as their reasons provided for holding different opinions. This central theme also included both positive and negative opinions of the utility of the tests for themselves and for other teachers. Within this theme, the data showed participants' identification of the problematic area of standardised literacy tests' educational uses, as the issues affected the classroom and the wider school setting. In addition, factors and influences beyond the classroom and broader school sites were incorporated in this central theme.

4.4.3 Theme 3: Pragmatics of standardised literacy tests.

The third theme, 'Pragmatics of standardised literacy tests' was made up of the re-examination and refinement of five categories: 'Classroom control' (interview category 1), 'Issues in test administration' (interview category 4), 'Broader test critiques' (interview category 5), 'Teacher effects' (interview category 6) and 'Pragmatics of standardised literacy testing administration, and administrative uses' (test session category 1). This central theme underpinned the data through those categories and reflected participant-perceived patterns and issues regarding how to actually administer and run the tests.

Participants' chosen approaches to testing, informed by participant-identified issues and influences within standardised literacy testing, were clarified as this theme emerged. One key pattern highlighted those participants' criticisms and critiques of testing that played a role in influencing the rules participants upheld and ignored. A second key pattern within this theme was patterns of classroom interaction and test administration, which were consequentially related to the participants' decisions regarding how to run test sessions.

4.4.4 Theme 4: Judgements on student communities, standards, attitudes.

The fourth theme, 'Judgements on student communities, standards, attitudes', incorporated the four related categories, 'Judgements of students' (interview category 3), 'Judgements on student communities and student catchment' (test session category 5), 'Judgements on student standards, learning levels' (test session category 6) and 'Judgements on students themselves: behaviour, discipline, self-control' (test session category 7). These categories all related through the clear central theme of 'judgements' concerning the school site's student cohort.

These student judgements were expressed in different ways. For instance, one pattern of the expression of student judgements showed participants' own perceptions and attitudes towards the behavioural and academic abilities that they considered characteristic of the student group and school. Judgemental attitudes towards the students, opinions expressed by participants and held by them or others, were shown through the data to be

expressed via explicit comparisons of their students against assumed characteristics of the student cohort at other schools. What participants presented as broader community judgements of the student group represented a sub-pattern of this theme. Participants' interpretations of their students against other students could be seen as an effect of those perceived community judgements. Another pattern within this central theme highlighted the various criteria against which the students were unfavourably assessed: behavioural norms; academic levels and abilities; and student community, such as perceived home and environmental difficulties.

4.4.5 Theme 5: Perceptions and behaviours of teaching.

The fifth theme, 'Perceptions and behaviours of teaching', is made up of the categories that demonstrate the central theme of teaching, in action and perception. The categories within this theme are 'Teacher effects' (interview category 6), 'Teacher as teacher' (interview category 7) and 'Direct, individual teacher-student interaction' (test session category 8). These three categories related to the overarching theme of perceptions and behaviours of teaching; the ways that participants approached, carried out and considered their teaching duties were identified as two main patterns.

The first pattern highlighted participant actions and behaviours while classroom teaching; this pattern subsumed the sub-pattern of teacher interaction with students while in the classroom teaching environment. This pattern of teaching actions and behaviours included instances and approaches to making decisions regarding the administration of learning topics, provision

of assistance to students, and the running and overseeing of activities. The second main pattern within the theme was that of participant regard to their effect on, or influence over, students. This pattern highlighted participant considerations of the affect their actions could have on students' learning, including student test performance. An emphasis that emerged from this theme pattern was the acknowledgement by participants of a multitude of decisions and interactions necessitated through their teaching and testing roles. Additionally, test administration forced participants to call upon their teaching skills and experiences, in making decisions and responding to situations.

4.5 Summary of Coding Analysis

This chapter presented the results of the analysis of the codes, categories and themes constructed through the initial, focused and axial coding stages. The various issues raised by the participants were reflected, and the discussion highlighted, the participants' patterns of response to the expectations and requirements of them as both teachers and test administrators. The coding allowed for the identification of various issues pertaining to participant opinions about the tests, the ways the tests were administered in the classrooms and their data used. Participants identified and perceived specific merits and flaws in standardised literacy test design, implementation and application.

The detail provided about the coding stages in this chapter showed that a number of codes constructed from both data types referred specifically to behaviours particular to test administration, participant responses to student approaches when undertaking the tests, instructions given for the students in

taking the tests, and teacher-student interaction during test sessions. Such issues and behaviours were further evident in the categories that were constructed through a further process of patterning of the initial codes. The categories highlighted, clarified and grouped these data patterns. Similarly, the themes from the axial coding showed overall patterns and drew attention to the ways that the codes and categories related. The initial, focused and axial coding of the test session transcript and interview data showed overlap, divergence and conflict in the themes across those data.

The results of these first three coding stages enabled five themes to take shape and facilitate this research focus on teachers' professional opinions of standardised literacy testing, and on processes and practices that can be considered to take place during teachers' administration of the tests. The two research questions, 'How do teachers perceive and interpret the value and validity of standardised literacy testing?' (research question 1) and 'What patterns of action and interaction characterise standardised literacy testing?' (research question 2), were addressed through the fourth coding stage, by which the themes were re-examined and the dominant discourses regarding the research questions and within the data patterns identified. The results of that process are presented in the following chapter.

All five themes enabled the research questions to be addressed by highlighting from the data how teachers' opinions of the tests' educational value and merit (theme 2, research question 1) were connected with their behaviour during testing (themes 1, 3 and 5, research question 2). Participants' judgements of the student group and school culture (theme 4) affected both the

participants' opinions of test value and validity for their students' and school's needs (research question 1) and shaped their actions and interactions in the classroom site of testing (research question 2).

The following chapter, Chapter Five: Research Questions and Themes, presents the results of the close examination of the five themes and the ways they addressed the two research questions. Limited boundaries between the themes and research questions will be imposed and, as already explained, the key data types (interview and testing session transcript data) are considered concurrently from this stage in the analysis. The themes synthesised the results of the coding of each data type, and therefore are used hereafter as conceptual and theoretical elements. The combined further analyses of the collected data types demonstrate the interconnections and overflow between and across all the collected data, that is, the key and supplementary data from field notes and questionnaires.

Chapter Five: Research Questions and Themes

Introduction

This chapter concludes the presentation of coding analysis results commenced in the previous chapter. The previous chapter, Chapter Four: Coding Analysis of Data, presented the results of the first three stages of the coding analysis, from which the five themes emerged; this chapter presents the results of the final coding stage, theoretical coding. The results of this theoretical coding stage build on, continue and conclude the previous coding stages to construct theory grounded in the data and identify the dominant discourses.

The research questions serve as a guide for the presentation of this final coding results stage by providing a clear structure through which the themes are framed. The results of the coding analysis (Charmaz, 2001, 2005, 2006) are presented through a clarification and exploration of the ways that they (particularly, the patterns noted across the five resultant themes) address the research questions.

This chapter moves beyond the previous chapter's noting of the various codes and categories that constitute the themes. This chapter contributes increasing depth and theoretical interpretation of the meanings within the data. The previous chapter described and exemplified the processes and results of the coding stages, to provide the researcher's mandate in identifying the most illustrative examples of the central themes located in the data. Examples drawn from the data, therefore, show variations within and across participants' arguments.

To reiterate the analysis detailed in the previous chapter, four stages of coding examined in increasing depth the content and meaning of the data. Three of these four stages were presented in Chapter Four. The first stage of the coding, initial coding, involved a close reading of the key data types to identify 'active' codes. The initial coding produced 44 such codes from the participant interview transcripts, and 39 codes from the testing session transcripts. The initial coding stage was followed by that of focused coding, through which the codes were grouped as patterns of similarity and effect into categories. The focused coding produced seven categories from the participant interview transcripts, and nine categories from the testing session transcripts. Once the categories had been constructed, the third stage, axial coding, broke down and refined the codes and categories to identify broader, theoretical concepts that became the themes, each of which are made up of various patterns and parameters of connection. The fourth stage, theoretical coding, presented in this chapter, identifies the overarching, core concepts, from which all other levels of coding can be related (Saldaña, 2009).

Until the axial coding stage, only the two key data types (the participant interview and testing session transcript data) had been analysed; as a result of this, only those two key data types were utilised in the previous chapter. These two key data types had been analysed separately, to ensure that the coding stages constructed grounded data reflective of the constitutive issues and patterns of each data type. The initial separation of data enabled the construction of data codes and categories that highlight, rather than skim over, the individual patterns and concepts characterising the different data types.

Separation of the key data types was no longer purposeful following the axial coding stage, as the data and the categories could then be synthesised to identify thematic and theoretical patterns broadly across the data. The researcher's field notes and questionnaires (supplementary data) strengthen this particular chapter, contextualising the themes and discourses identified through the coding analysis, and including the researcher's 'voice' on the initial interpretation of phenomena. The researcher's field notes and questionnaires are included in this chapter to compare and contextualise all collected data types against each other and against the themes and emergent discourses from the key data coding analysis. The synthesis of data types recognises this difference by identifying the strongest patterns and their patterning, parameters and properties.

Theoretical coding is so-termed because it is from this stage that the analysis advances or moves into the realm of theory and begins to challenge, extend or generate new ways of thinking about this topic. Glaser (1978) wrote that theoretical codes (in this study, the dominant discourses) conceptualise "how the substantive codes may relate to each other as hypotheses to be integrated into a theory" (p. 72, in Charmaz, 2006, p. 63). As such, they "not only conceptualize how your substantive codes are related, but also move your analytic story in a theoretical direction" (Charmaz, p. 63). By theoretically coding the data and addressing the research questions, grounded theory is begun to be generated, that is, theory grounded in the data. Theoretical coding is conducted in this chapter by engaging more reflexively and critically with the themes than during the axial coding stage. The theoretical coding must "fit

your data and substantive analysis” (Charmaz, 2010, p. 63), and the lens of the research questions enables a clarification of the patterns and concepts in the data by way of utilising the research questions as a guide for presentation.

As stated in the previous chapter, the five themes identified through the coding analysis were:

***Theme 1. Maintenance of control, behaviour
management, order***

***Theme 2. Utility of standardised literacy tests as
educational/learning tools***

Theme 3. Pragmatics of standardised literacy tests

***Theme 4. Judgements on student communities,
standards, attitudes***

Theme 5. Perceptions and behaviours of teaching.

Figure 5-1. *Themes.*

The ways in which these themes address the two research questions are detailed and discussed throughout this chapter, as patterns are restructured for clarification of meaning and data types are blurred to present a synthesised and critical response to the research questions. This chapter incorporates examples from all four data types through sections dedicated to each research question.

This chapter does not present an exhaustive exploration of all elements of each theme or data type, not least because the theoretical positioning that frames this research problematises the idea that an entire ‘truth’ can ever fully be recognised by a researcher. Rather, the chapter explores those elements

within each theme and data type selected by the researcher for their illustrative and explanatory power, because the data highlight the participants' special focus on particular underlying issues. Examples from the data also show variations within and across participants' arguments. The previous chapter presented illustrative quotes from the data to provide the researcher's mandate in identifying the most illustrative examples of the patterns and concepts. In this chapter too emphasis in quotes and data extracts represents emphasis in original. This means emphasis identified in the transcript, by speaker (in the case of audio recorded transcripts) or by the author (in questionnaire responses and field notes).

This chapter is structured in three sections. The first section, 5.1 Research Question 1, presents the results that address the question 'How do teachers perceive and interpret the value and validity of standardised literacy testing?' The second section, 5.2 Research Question 2, presents the results that address the question 'What patterns of action and interaction characterise standardised literacy testing?' Those two sections break down, deconstruct and blend the themes and data types. The third section, 5.3 Results, concludes the chapter with a brief reiteration of the core, dominant concepts identified from the data through the coding analysis and from addressing the research questions. This final section leads the coding analysis, as an introduction, into the discourse analysis, which is the focus of the following chapter.

5.1 Research Question 1

Research question 1, ‘In what ways do teachers account for the value and validity of standardised literacy testing?’, was primarily addressed through the patterns constituting theme 2, ‘Utility of standardised literacy tests as educational/learning tools’. The patterns within theme 4, ‘Judgements on student communities, standards, attitudes’, will also be incorporated as and where relevant. These patterns across the data demonstrated the variety of participants’ opinions regarding the value of standardised literacy testing and validity of the tests. In the context of research question 1, ‘value’ refers to importance and worth attributed to testing and to test data. This particularly refers to the ways in which tests are considered valuable and the uses they are put to. ‘Validity’ refers to the accuracy of these tests’ results as indicators of student literacy learning and abilities.

The participants discussed three main values or uses made of standardised literacy testing and test data. These were: literacy diagnoses (5.1.1); school-wide literacy programme planning, largely as a result of test timeframe issues (5.1.2); and political purposes (5.1.3). Participants’ varying opinions about these values and uses made of the tests and testing are detailed, with illustrative extracts from the data.

In addition to the main ways that standardised literacy tests and testing were identified as valuable or valid for use, participants also presented three key ways in which the value and validity of standardised literacy testing could be affected or undermined. The first of these is test timeframe issues, explored through the broader topic of school-wide literacy programme planning (5.1.2),

the latter largely being a value made of test data because of this issue. The second of these issues is the depth of detail provided in test data (5.1.4). The third is an overarching pattern of influences upon test sessions (5.1.5).

5.1.1 Literacy diagnoses: “pretty parallel” / “helping the right ones”.

The dominant value the participants presented of standardised literacy testing for themselves and their students was as a tool for literacy diagnosis, a tool that validated their judgements of students. Test data were valuable when they indicated areas of student literacy strength and weakness, and as a check or comparison against students’ in-school ratings and assessments. Participants’ statements about this value of standardised literacy testing and test data further demonstrate that they differed in their opinions, but that they appeared to consider testing educationally valuable as a complement to their reflexive judgement.

The first key pattern identified here is the participants’ ready noting of the particular value for teachers of testing for literacy diagnosing and validation purposes. The extracts below are examples of this pattern.

TT: ... we use them as a diagnostic tool once the tests have been completed, in that we might look and see that individual X has a particular weakness that we may not have noticed. Um, the other thing we often use them for is that the results are used to confirm or to make us question the ratings we are currently giving [the] students. (TT, Int. 1, p. 1, L 13-17)

TT: But certainly it’s the sort of thing that you use to, A, validate your own assessments, always makes you feel good if you see that they’ve come up with the same answer that you’ve come up with, and B, if somebody is way out of the range, then you would spend

a bit of time trying to work out why that might have happened, or reassessing your own view of that individual. (TT, Int. 3, p. 1, L 30 – p. 2, L 4)

T2: And I suppose the good thing is, that all the kids in my home class who are grade sevens [who] are in the reading tutoring programme, are the ones who showed up as the ones who really needed it, so there's a good correlation, so that was good, that you're helping the right ones. (T2, Int. 1, p. 4, L 13-16)

T2: ... the one thing I always try and say is, 'how well you do on this does matter'. Not just because you know someone else somewhere else is going to make up all those pretty graphs and do all those sorts of things, but simply because we really want to know whether you're progressing or if there are any problems appearing that we might not have known about, I'd say that 'it's for your benefit as much as for anyone else's', I always try and put it that way ... I just say to them, you know, 'we're really interested to see where you've come since grade seven, and that's two whole years and we'd be hoping that there'd be really big improvements from everybody' (T2, Int. 1, p. 10, L 10-14)

The above extracts highlight these participants' opinions that these tests are valuable even in their validation of teachers' judgements and decisions, a theme continued in the dialogue of other participants. Participants presented standardised literacy testing as valuable for teachers as an identification of student literacy progress and areas of student performance that needed teachers' diagnostic attention. The patterns evident in the above extracts is also evident in those participants' questionnaire responses. On his questionnaire, TT noted that testing sometimes highlights differences with or validates his own assessments of students (TT, Questionnaire, Q3a, Q3d). T2 noted that testing facilitates her "Development of individual programme/extension work" (T2, Questionnaire, Q3a), and TT noted that test

“results may influence my choices of subject matter” (TT, Questionnaire, Q3c). The student questionnaires, too, indicate the observed students’ overwhelming affirmation that they considered the tests positively affected their learning (see Appendix C).

The extracts demonstrate TT and T2’s opinion that, in serving as a check against students’ subject results, test results enabled and justified decisions about literacy ratings, diagnosing and additional assistance provided to particular students. The participants’ use of test results to check their assessments of students was communicated as a source of educational value for teachers within standardised literacy testing. In this way, participants clearly presented that it was necessary for standardised literacy testing to be ‘valid’, as tests could present a good, valid, “correlation” with students’ non-test literacy performances. The extracts above, further, showed that test results could be valuable for ‘future’ purposes. Firstly, standardised literacy tests could shape students’ educational futures by allowing teachers to predict or validate student literacy progress and academic placement; and secondly, the tests over time clarified students’ progression of literacy ability. Such comments suggested a belief in the validity of standardised literacy testing.

The second key pattern in this aspect of testing is that of participants’ identification that the tests are further valuable, particularly in the words of TT and T1, as helping to illuminate areas of literacy weakness that students could otherwise hide during internal assessment tasks but not during standardised literacy testing. The extracts below present examples of this pattern in TT and T1’s own words.

TT: The place where literacy testing is really beneficial, the place where literacy testing is really valuable is it will pick up those individuals who are good at hiding the fact that they're not good readers, and at this stage, reading out loud doesn't happen all that often, and often they can hide the fact that they don't read very well. They can hide it in all sorts of ways, and a literacy test tends to make it sort of stand out, because their comprehension answers will be very – they'll get a very low score for comprehension, and that should ring alarm bells about whether or not this person is really understanding what they're reading. (TT, Int. 1, p. 3, L 28-p. 4, L 4)

T1: I think some of them were very much aware that their inability to perform in literacy work would be out there, in stark reality, to see. You can't hide from that sort of thing, and I think they were quite aware of that in certain times, in many classroom situations they could cover it, could hide it, avoid it, copy, cut and paste from something. But you got a pen and a pencil, you've got a blank sheet and your name on the top and you're on your own. (T1, Int.1, p. 2, L 31- p. 3, L 5)

TT focused on students' reading abilities, given as an example of an element addressed by the tests. TT interpreted a "very low score for comprehension" as a warning that the student was unable to read for meaning, similar to the second use of standardised literacy test results in the earlier quote from TT (TT, Int. 3, pp. 1-2). T1 echoed TT's sentiments that the tests highlight what could otherwise be missed by teachers, also identifying a major aspect of the tests that could be an indicator of the value of standardised literacy testing for teachers. T1 indicated his opinion that students could hide literacy issues through use of digital technologies in learning and assessment tasks. For the student lacking confidence to academically perform when under pressure, unable to "hide" or "avoid" their literacy issues, TT and T1 opined that these tests forced a recognition of areas needing diagnostic attention.

The third pattern in the broader theme of literacy diagnosis is participant response to literacy information communicated in test data. TT, T1 and P expressed surprise while commenting that test results served as a check against internal student assessment ratings. The extracts below example this pattern.

TT: But incredibly, the standards that they nominate for most of the kids are not too different to what we would come up with anyway. ... Which tells us that fools never differ, perhaps. (TT, Int. 3, p. 1, L 17-20)

T1: I looked at the sheet the other day when we got them, and I looked through all the kids I taught in grade 7 and 9, and there weren't any surprises, in my eyes at least. From my observations, and from my watching in the classroom to what I saw there. They were pretty parallel. (T1, Int. 1, p. 5, L 23-26)

P: It always is a surprise to me that the results actually come out as well as close to where – well, with my English class, that's about where I'd been assessing them. (P, Int. 1, p. 5, L 13-15)

TT, T1 and P articulated, with specific reference to “surprise”, that the test results reflected their expectations and own assessments of their students' literacy skills. While the P expressed surprise at the similarity of test and internal results, implying his expectation of find some differences each year, T1 stated that the test results contained no surprises (“in my eyes at least”), meaning no differences. T1 implied an expectation that the tests would provide an ‘accurate’ reflection of what he himself had identified.

5.1.2 School-wide literacy programme planning and test timeframe issues: “how we’re progressing” / “way too late”.

The dominant pattern across this theme identified participants’ explicit sense of frustration at what they argued was the lateness in the school year that tests results were returned to schools, with the implication being that the value of these tests for teachers and those at the school site of testing was negatively affected. The following extracts are examples of participants’ statements on this issue.

TT: One of the other criticisms I’ve written on that survey is that the real problem is that the test results are not appearing in the school until the year is nearly over. ... Because of the demands of our reporting schedule, it all has to be done well before the end of the year, and so the results are going to come back after I’ve written the reports. So not only is it no use in terms of assessing the child, it’s absolutely no use as a diagnostic tool. Because anything that does jump up from the results is revealed to you way too late in the year. We need to know this stuff in the beginning of Term 1. And yes, it could be used in the following year, but the nature of the organisation of schools is such that often classes get passed from one individual to another, and that stuff sort of fades into the ... Oh, the data could be really useful. But we don’t see it. (TT, Int. 3, p. 3, L 17 - p. 4, L 14)

T2: You get them in time to send them home to the parents. (T2, Int. 1, p. 8, L 1)

Participants claimed that they were unable to utilise test data in the same school year in which they were generated, instead having to effectively ‘make do’ without being able to use this valuable information. TT was particularly explicit in his frustration at the late return of test results and what this late return of results meant for their value.

Although TT implied that some fault lay with school-level demands and issues, and apart from the issue of why teachers apparently do not use previous years' test data, TT clearly expressed his frustrations that the delay undermined the value of standardised literacy testing. He provided reasons why the tests were less useful, by stating that such a late return means the teachers received potentially valuable information "way too late in the year" and that the reality of the school setting was such that test data were often not used even in the next school year by successive teachers. This issue was evident in TT's questionnaire responses, with his identical comment regarding timing of data return to schools, but to which he also added the suggestion that "testing should be done much earlier" (TT, Questionnaire, Q5). The student questionnaire too included one student suggestion, "give the teachers a card on how they went for their class to see where students are at" (see Appendix C), which indicates student concern about their progress and desire for information about their test progress.

Another repercussion for the value of the testing for teachers, resultant of this pattern of time-frame issues, is represented by TT and P as limited use for individual students but instead school-wide indicators.

- The results come out so late in the year that they can only be used mainly for programme planning than for working with students individually about problem areas. (P, Pre-Int. 1, Field Notes, p. 1)
- Feedback valuable in assessing curriculum direction (TT, Questionnaire Q6c)

TT: But, in general terms, they are the sorts of things that we use more as a school-wide indicator of how we're progressing, rather than as an individual diagnostic activity. (TT, Int. 1, p. 1, L 22-24)

Both P and TT identified that standardised literacy test data were used for broader school purposes, namely to contribute to whole school-level literacy diagnosis and programme planning. TT above claimed that test data were often used for this school-level purpose. P presented his claim about broader school-level uses made of standardised literacy testing data as a negative, caused by insufficient time in the school year with which to make any other real value of the tests. Nonetheless, as already illustrated, both TT and P also claimed to use test data to engage in some comparison, check and diagnosis activities.

The third identified pattern in this theme is participants' expressed opinions of the uses and values that testing could represent, should the time frame issues be resolved. The following extracts provide examples of this pattern. To contextualise the first extract given below, T2 had been asked if an earlier return of test results to schools would allow the teachers to find test data more helpful or valuable.

T2: I'm sure you would. Because even though I said the thing about the reading programme and the students that most need it did seem to get fairly well correlated, I still think that if you knew that – this student who apparently spells quite well actually isn't a really good reader, or that this one who can read well and can spell well really has to develop in the writing skills area, then you could do something about it ... I actually read through their writing tasks before we actually – I managed to squeeze in pretty well all of them and made a few notes before we actually got – well, long before we got the information back, because I thought it was useful to see how well they could write in a situation where they had a set time, and how well the stimulus picture worked for suggesting ideas and, the idea of planning it (T2, Int. 1, p. 7, L 9-19)

R: If the results came back – or if you could do them really early in the year and the results came back really quickly – what would you actually be able to use it for? Do you think that you'd be able to provide more focused support for the students –

TT: Absolutely. It would allow you to individualise the curriculum that you provide, because it would give you quite a bit of guidance about skill levels, or more importantly skill absences, for particular individuals. And often, particularly early in the year if you've got a new class, it takes quite a deal of time to actually understand exactly where each individual is at. So a test like this could be really valuable in pointing out glaring areas of concern for particular individuals.(TT, Int. 3, p. 7, L 1-9)

TT, T2 and P agreed that tests would be more valuable (for teachers, students, and schools) if test results were returned to schools earlier in the school year.

T2 admitted to having resorted to examining and using the information she gained from her students' test responses before submitting those responses for computing and processing. T2's comments show that she considered standardised literacy test data to be valuable, as otherwise she would have seen no purpose in using what her students had produced, and what the test tasks had highlighted, before the computed results were returned later in the year.

By indicating “glaring areas of concern for particular individuals”, the test had major value to TT. This was particularly the case early in the school year before the teacher had much exposure to students' “skill levels” and “skills absences”, which in turn would enable a more “individualise[d]” curriculum. As already noted, P raised this issue as an undermining of test value due to test timeframes as early as the first meeting with the researcher, in his claim (recorded in the researcher's field notes journal) that test results were returned so late in the year that the value of the data was minimised or negated in its value for teachers in individual student diagnosis. TT, T2 and P stated

that standardised literacy test data were useful for teachers, but that issues caused by testing timeframes and the protracted return of results to their school meant that standardised literacy testing had reduced value for them. To contextualise this timeframe issue, the standardised literacy testing sessions ran over two weeks in July/August, and the results were returned to the P some three months later, in November of that year.

This issue of delay leads to the conclusion that the value of standardised literacy testing for these participants was undermined even before the testing took place. Given that the results were returned so late in the year that teachers would already have identified their students' literacy skill "levels" and "absences", it could be argued that the tests were too late altogether. TT and T2 expressed contradictory and mixed opinions on this point. TT stated, as already noted, that the tests are "no use in terms of assessing the child" or "as a diagnostic tool" (TT, Int. 3, p. 3), but also stated that the tests can "ring alarm bells" (TT, Int. 1, p. 4) and guide the teacher to "[reassess] your own view of" students (TT, Int. 3, p. 2). T2 also identified such benefits and values of standardised literacy testing, as already noted.

5.1.3 Political use made of test data: "more of an imposition, than anything".

In addition to any educational value of testing, these participants identified that the data were politically valuable. The first pattern here is political and accountability value of standardised literacy testing. TT, T1 and T2 presented the issue of political use made of test data by politicians and governments as a source of frustration and as undermining the educational value of standardised

literacy testing. Participants' discussions of this issue suggested that discontent over politicisation of testing and test data was widespread. The extracts below are among the examples of this pattern through the data.

TT: It's probably a bit politically incorrect to say so, but I think there's a feeling amongst a lot of people that these sorts of tests are run so that politicians can claim that they are monitoring literacy levels and they're making schools accountable and use them as evidence in their political manoeuvrings. (TT, Int. 3, p. 4, L 6-9)

TT: I suspect, people would be much happier to go through with the process if they thought that the end result was something that was going to be of value to them. Most of us, I suspect, believe that the testing is more of an imposition, than anything, and as we've already discussed, the political ramifications, et cetera, often make it difficult to feel all that enthused about it. (TT, Int. 3, p. 6, L 18-23)

T1: One wonders about the political mileage that is gained. ... I would imagine that it's a marketing exercise. And a political exercise in justifying funding regimes, justifying staffing ratios or whatever, I'd imagine they're for. But then, it's used in the other way, isn't it, by the Federal Government – 'you've been a very naughty school, your literacy levels are lousy or low, so you haven't done a very good job as teachers – so we're going to cut your funding'. Which I think is a counter-productive argument. (T1, Int. 1, p. 6, L 5-31)

T2: Uh, think it's – they certainly tend to serve a political purpose, because every time figures come out there's always some politician jumping up and down, and saying, you know, schools in this State are X points behind, blah blah, and oh god, for heaven's sake, they love to take it all totally out of context and not see 'either this child has come forward this far from the last time he was tested', or they're not interested in that. Their ideas of indicators of success in schools are just – you know, so I don't like to see it used as a political football or as a political threat, which, of course, they do. (T2, Int. 1, p. 8, L 17-23)

TT's scepticism and frustration regarding the political use he believed was made of test data highlighted the difficulties that he implied obstructed teacher enthusiasm about testing. The above examples highlight the participants' frustrations at politicisation of testing, which TT posited was experienced by other teachers as well. TT indicated his unhappiness about testing, given a lack of direct value from testing for teachers, and as such the testing somewhat represented for these participants "an imposition", compounded by "the political ramifications" and decontextualisation of the testing. TT's questionnaire responses too indicate his frustrations and apparent conflict about this point, with his statement that tests "are a useful tool. With accountability being central to our world. They are valuable" (TT, Questionnaire, Q7).

The issues of politicisation of testing and political values and uses made of test data were particularly problematic for T2. T2 expressed her strong view that standardised literacy testing was educationally valuable and, therefore, she disapproved of test results being used for other (and perhaps less 'valid') outcomes. The terms "political mileage" and "political exercise" of standardised literacy testing discussed by T1 relate to T2's mention of "political purpose", "political football" and "political threat", and with TT's description "political manoeuvrings".

The second key pattern regarding this way in which testing is considered valuable pertains to consequences of such political and accountability uses. T2's already-presented reference to State and school comparisons ("schools in this State are X points behind", T2, Int. 1, p. 8) were similar to sentiments expressed in the following extracts regarding one

potential repercussion of a political rather than educational use being made of standardised literacy testing.

TT: Even *more* ridiculous is the issue that the current Federal Government are raising, where they're talking about monitoring the performance of teachers, and paying them accordingly. ... Yeah, so this would be the sort of thing that they could, theoretically, do, to do that, and of course if you work here, your results are not going to be as good as the people who work at [State co-educational secondary school], because of the clientele that you deal with. So it's just a mine-field. That sort of policy is so simplistic, as to be laughable. (TT, Int. 3, p. 4, L 20-27)

T2: The idea of league tables, which they [the broader community] just love the idea of league tables, I mean, comparing us with [State all-girls secondary school] is like comparing a cheese stick and a carrot, I mean, they're not the same thing at all. Look at the demographic and the sorts of kids who might be expected to come here but for many reasons don't, we've got a huge range of, you know, ability in our school and when you try and even it all out to see where a school is at, I think it becomes pretty meaningless, actually. (T2, Int. 1, p. 11, L 21-26)

- Can be good or bad, if used to 'rank' schools – or punish schools which fail to meet benchmarks. (T2, Questionnaire, Q6c)

TT expressed concern that judgements of schools, teachers and students “could, theoretically” be made on such a “simplistic” basis. TT was consistent in his argument regarding the issue of political value found in standardised literacy testing results in judging teachers and students based on school comparisons and T2's statements reveal hesitation and possible defensiveness, expanding on factors negating school comparisons. According to T2, controlling for differences served no helpful purpose either, making even those comparisons “pretty meaningless”.

5.1.4 Detail provided in test data: “that’s about the best we can expect”

The amount of detail regarding students’ performance in particular tasks and the overall testing requirements was a further key issue that participants indicated affected the value of standardised literacy testing. Within this theme, the identifiable patterns pertain to comparative detail and clarity of data presentation, as well as the data provided to schools and students when data were returned

The first pattern, exemplified through the extracts below, demonstrated that the detail provided following testing processes was important for participants in making use of, and perceiving value in, test data.

- The results report on reading, writing, overall literacy, and overall numeracy. They used to give more detail, breaking down into strands (those in the Statement and Profiles and other English curricula and syllabi), which was more helpful. (P, Pre-Int. 1, Field Notes, p. 1)
- P: The problem was, some of the stuff that’s ... has been presented in a different format, uh, than normal. We usually have coloured graphs and bars, a single graph, a single column, single line bars with gradations in colour. This year they’re doing it somewhat differently, as you can see where the score – it’s the same way as the individual ... really. (P, Int. 1, p. 2, L 24-28)
- P: Now, as far as the end of the year, kids’ll get this, there is an explanatory note at the top, isn’t there – yep. They’re not that user-friendly ... But that’s about the best we can expect. (P, Int. 1, p. 5, L 21-24)
- T2: I guess that’s, I mean if you look at the tests and you see that there’s an improvement, then that’s something, but if you’re just looking at you know ‘this is a child in grade seven who is here, children in grade seven really are supposed to be somewhere

between here and here', that doesn't really mean much for that child. (T2, Int. 1, p. 3, L 17-20)

T2: I mean, there are – one lot with all these graphs and lines and things, and [P] just says 'look don't worry about that bit, I'll need some professional advice on how to actually interpret these', like the overall figures and things. (T2, Int. 1, p. 4, L 24-26)

P was openly critical of the detail provided in test data and results for schools, especially when compared with previous years' reporting and presentation. Detail in test data and results was significant given the clear relevance of test results for curriculum purposes. Schools were permitted to apply information directly to literacy learning programmes, however the lack of explicitly and deliberately relevant detail frustrated those school developments. P seemed to expect little from the test data and result information returned to schools because of a decrease he noted in the detail provided, compared to previous years.

P further indicated that current approaches to data presentation and depth of detail were not ideal, noting that those at the school-level needed easily interpretable test results. P's communication of low expectations showed that he considered teachers, principals, parents and students received less than ideal information regarding students' literacy standards. T2's questioning of the value of standardised literacy testing for students themselves when given only basic comparison of performance against a benchmark extended P's claim about the usefulness of particular approaches to data presentation and depth or type of detail. In addition to the style in which data and detail were communicated, T2's comment here indicates that she held comparative test results – in which students' achievement levels are compared

to their peers’ – to be less valuable than more detailed reasoning of precisely the skills in which a student demonstrates excellence or lacking.

The detail and information provided regarding test results also affected parents. This was an issue that TT and T2 discussed, as the following extracts show.

TT: The only results our parents get – might be different at other schools, but in our school, the only results the parents get are subject-related results linked to the Essential Learnings. I don’t know, even, what the legalities are of releasing the results. Certainly it’s not something I would do without consulting with the hierarchy in the school. If a parent called me, and said, ‘how did my child go in the test’, I don’t even know what the policy is on releasing those results. (TT, Int. 3, p. 3, L 2-8)

T2: I never know what parents think, I’m never sure about what parents think about the tests. (T2, Int. 1, p. 12, L 22-23)

TT’s expressed uncertainty about the usefulness and value of test information provided to parents was echoed in T2’s expressed her uncertainty regarding parents’ interpretations and opinions of standardised literacy testing. The above opinions were also expressed succinctly on their questionnaires: “Have no access to results”, “I’m never sure what parents think about tests” (TT and T2, Questionnaire, Q 6b).

5.1.5 Influences upon test sessions: “that’s what happens. In the test situation”.

All participants considered that influences on test sessions strongly affected the ways that students approached their standardised literacy testing tasks, including students’ test responses. Student performance during test sessions – and, by extension, their success or otherwise on test tasks – was influenced by

such factors as student behaviour, concentration levels, broader community problems, class size and background of students.

The major pattern across this data theme is that of participants' acknowledgement and identification of just such influences upon test sessions. The extracts below are examples of this pattern.

- TT: ... there are issues to do with extraneous problems that might be happening in the community, which can affect the reactions of individuals on any particular day (TT, Int. 1, p. 2, L 11-13)
- TT: So we've got three kids who are functionally illiterate, and they're gonna take a minute before they realise they can't do it, and then we're going to have to keep them quiet whilst the others get on with it. (Year Seven Reading Test, p. 1, L 9-11)
- T2: ... what's the best time of day to do it – obviously I don't want to give it to them in an afternoon lesson, because they won't do as well, so I never would, it just wouldn't be fair. (T2, Int. 1, p. 6, L 16-18)
- P: And the ... at the end which is saying that – that ... there, about the progressions statement, and progressions, and standards, is rather interesting. So it just says that they may not have shown, if you like, competence towards it. And it's significant, I've just picked up a grade nine class at the start of this term and we're doing some extension work, and one of the students, whose first pieces of work I graded at 5 Middle, he's only been assessed at a 4 Upper, so in terms of those kids at the upper end of the scale, it's not bad, but it depends on what sort of day they're having the day of the test, because there are other results I've gotten which are below and above, so it's – OER, to their credit – oh, well, yeah, to their credit – do make the point this is one test. They do make the point. (P, Int. 1, p. 1, L 17-26)
- P: [7FS1], interestingly – I reckon she didn't put any time into her score, because [7FS1]'s not dumb. Into her performance.
- R: No. I was in there when she was doing it.
- P: Were you? Yeah.

- R: I think she was getting really annoyed because –
- P: Yep.
- R: – because the boys in her class seemed to be getting a teacher's aide to work with them and she didn't want to do it and she got angry that they were allowed to – she just stopped.
- P: And that's what happens. In the test situation. If we look at the low scores in writing, we'll probably find more. (P, Int. 1, p. 3, L 30- p. 4, L 9)

Participants opined that influences upon students might be beyond teacher or school control, although teachers need to respond appropriately to administer classroom tasks. T2's obvious concerns that testing benefits could be jeopardised by unfairness echoed the similar fairness issue explored in 5.2.1.2, and was emphasised by her evident belief that these tests were helpful and beneficial. Her reliance on student test data to be valid, fair and accurate representations of student skills in order for her to be able or willing to make use of test data and results meant that fairness was a genuine concern.

Disruptive student behaviour, some of which was predicted even prior to the test sessions, was presented as influential on those students' and their peers' test "performance", as well as other various influences. P's reaction at the disclaimer accompanying the returned test results, that "one test" on its own cannot categorically show a student's literacy levels but rather indicate what the student's literacy progression might be, appeared to echo participants' arguments. Despite a degree of reluctance to pay credit to test result information, and ready acknowledgement of influences upon test results, participants appreciated the support for their judgements in the form of the (limited) test results.

The second pattern within this theme, an extension of the first, presents participants' perceptions about test value and usefulness as affected by influences upon test sessions. TT and T2 strongly stated that test results were problematic. Although acknowledging various benefits and values of testing, these two participants suggested that other evaluations were more useful for teachers. The following extracts are examples of this pattern, and highlight the importance of appreciating these limitations upon testing.

TT: Obviously there are going to be statistical variations where people have a bad day or whatever it is, and that one test can't be used as a perfect indicator of what an individual is capable of. (TT, Int. 3, p. 1, L 28-30)

T2: Course I suppose my real thing about the tests is that they show certain things, but they're not the only measure of success. And you don't really want the students or their parents or the wider community to think that they are the only measure of a student's success. Or a school's success, as well. (T2, Int. 1, p. 11, L 11-15)

This pattern indicated that a single test cannot adequately reflect student abilities. TT's point that a single, summative assessment of student performance on one test could be affected by any number of influences, and therefore deny a broader view of student abilities, raised issues about measurements per se. T2, too, although not claiming that the test results were invalid or lacking in value – her opinion on the value and validity of testing has already been presented – did present her criticism that these tests would be over-valued, used to judge student and school success. TT's explicit reference to “a bad day” is echoed by T2's noting that students “Can be ‘tested out’; may not have a good day!” (T2, Questionnaire Q6). One student identified this as a genuine concern, with the note on a questionnaire that “[I] dont like it when

you have like 3 in a day it gets stress full when you try your best” (see Appendix C).

5.1.5 Summary of research question 1.

Addressing research question 1 through an examination of the patterns that constitute themes 2 and 4 identifies and considers the key similarities and differences across participant perspectives regarding test value and validity. Participants considered, and accounted for, the value and validity of standardised literacy testing in a variety of ways, adopting contrastingly positive and critical stances. Participants identified factors that influence test responses, and their concerns about affecting – through action or inaction – the validity and accuracy of results. Participants’ valuing of testing and test data as a validation of their teaching judgements was noted across the data, as were their frustrations and concerns about uses made of testing and test data, such as political point-scoring, teacher professionalisation and unhelpful school comparisons, of which the participants did not approve.

5.2 Research Question 2

Research question 2, ‘What patterns of action and interaction characterise the classroom site of standardised literacy testing?’, is primarily addressed through theme 1, ‘Maintenance of control, behaviour management, order’, theme 3, ‘Pragmatics of standardised literacy tests’, and theme 5, ‘Perceptions and behaviours of teaching’. As with research question 1, the patterns within theme 4, ‘Judgements on student communities, standards, attitudes’, are also incorporated as and where relevant. The patterns across the data that constitute

these four themes demonstrate that teachers and students engaged in a variety of behaviours and interacted in various ways during classroom test sessions.

In the context of research question 2, ‘action’ refers to the acts and processes employed by the test actors during test sessions, in order to administer and execute testing tasks. ‘Interaction’ refers to the interplay between test actors during test sessions, such as communication interchanges. These two are closely related.

The testing session transcripts and test session field notes highlight the patterns of teacher-student action and interaction through the researcher’s direct recording and observation. The participant interviews allow participants’ voices to be considered alongside TT’s approach to classroom test administration observed by the researcher and recorded in field notes. Considering researcher recordings and observations and participants’ voices from interviews alongside each other answers questions raised in the data about patterns of action and interaction, such as what could be seen as happening and how participants explain and provide reasons for their classroom approach.

Within these themes, four overarching patterns of test session action and interaction between teachers and students were identified: firstly, clarification and affirmation of test-related rules, instructions and expectations (5.2.1); secondly, an overarching pattern of responses to student questions and concerns over testing (5.2.2); and thirdly, an overarching pattern pertaining to teacher responses to students’ negativity to the tests and testing (5.2.3).

5.2.1 Clarification and affirmation of test rules, instructions and expectations.

Much of the interaction observed between TT and his students during the testing situation related directly to the tests themselves, particularly to the rules and expectations of the test sessions. In this sense, TT articulated expectations of students in the test situation, and his engagement in particular actions and interactions indicated what he perceived were the expectations upon him as test administrator.

The nature of test administration also meant that, during test sessions, teacher-student interaction was often directed at issues of maintaining order in the classroom. Within the pattern of instructions, rules and expectations regarding test session administration, TT and his students interacted in response to student acts of noncompliance with test rules. This pattern of behaviour-related action and interaction includes clarification and reinforcement of set or accepted behavioural rules. At its most extreme, this latter pattern of teacher-student actions and interactions regarding instructions, rules and expectations took the form of TT relocating students within the classroom, threats of student removal from the classroom, incident escalation and referral to senior school staff, and TT's refusal to engage with the noncompliant students. These elements of the pattern are explored in this section.

These two key patterns regarding teacher-student interactions over test rules, instructions and expectations identifiable through the data are clarification (5.2.1.1) and affirmation (5.2.1.2).

5.2.1.1 **Clarification: “All you need to do at this stage”.**

Within the general action and interaction relationship, there is a clear pattern of clarification of test rules, instructions and expectations. At the beginning of each test session, TT read to students the instructions for test sessions provided in the teacher booklets for just this purpose. In addition to this initial clarification and explanation of testing rules and expectations, during sessions, students queried what was to take place, and TT was frequently required to explain and reiterate rules already dictated to students.

In clarifying test rules, instructions and expectations, TT frequently explicitly referred to set guidelines and instructions in his interactions with students prior to and during test sessions. The following extracts are examples of this very common theme, and are all taken from the very first observed test.

- 9 am – after brief preparatory talk, teacher begins with the instructions written inside the test booklets. (Year Nine Writing Test [session 1 of 2], Field Notes, p. 7)
- TT: All you need to do at this stage is listen, and then we'll go through the words you need to spell. (Year Nine Spelling Test, p. 9, L 21-22)
- TT: Oh, I'm sorry, I got sidetracked by those people, I need to read you some instructions. (TT, Year Nine Writing Test [session 1 of 2], p. 5, L 13-14)

The observed test session data show that TT clearly articulated test instructions and rules to students, as part of the testing process. The data illustrate that, at the beginning of test sessions, the whole class group was read the instructions that would govern the session, and during test sessions, his focus on ensuring student awareness of rules suggested his concern with providing students with

the set instructions and information. TT attempted to ensure all students received proper and equal instruction prior to beginning their testing tasks.

The second theme within this pattern pertains to fielding student queries during test sessions. In properly explaining test tasks to all students, the data demonstrated that TT had to field student queries that interrupted the set order of student test information, and the observed students often requested information additional to that already provided. Particularly at the beginning of the sessions, students questioned TT as to what was to happen during the tests. The following extracts are among the many examples of this pattern.

S: Do we have to do the answers in pen or pencil?

TT: I will explain all that in a moment. (Year Nine Writing Test [session 1 of 2], p. 2, L 15-16)

S: Do we have to fill in all of the lines?

TT: No, you don't have to fill in all of the lines. As much as you can, to tell the story. (Year Seven Writing Test, p. 7, L 1-3)

S: But I don't know what to put in these boxes!

TT: I'll tell you what to put in the boxes. I'll call out the words. And all you have to do is work out how to spell 'em. Or do your best at spelling them. (Year Seven Spelling Test, p. 3, L 23-25)

TT often had to choose between which of many students' requests to respond to, and similarly had to decide how to respond to students' requests. Further, instances in which students queried rules and instructions that TT was allowed to provide to students demonstrate that, in order to minimise student stress and get through the testing sessions, students and teacher interacted in such a way as to clarify the rules and expectations that governed their testing. The above extracts also show that TT provided encouragement to students while answering their questions.

Teacher-student interactions also suggested that, when in a position of clarifying test rules, instructions and expectations, TT was at times unsure of correct or appropriate test procedures. TT's test administration indicated that test rules and expectations were not always explicated for either teachers or students. TT put some emphasis upon the students as well as upon himself, as administrator. TT's focus on what he and his students were required to do indicated his awareness that appropriate direction was necessary for students to understand the task, and also suggest his distancing of himself from such directions, which he emphasised as requirements of test tasks (not stemming from him) and upon teachers and students.

The third key pattern in the broad theme of teacher clarification of test rules, clarification and expectations demonstrated the role of behaviour management in such instances. The following extracts example that, in addition to his attention to informing students of and clarifying the test requirements and expectations, TT's test administration approach showed that he prioritised behaviour management during test organisation and administration.

TT: Just – let me get everyone sorted out, get everyone started and then we'll explain what those who might've finished ... [9FS5] ... don't start writing just yet please, till you know what we're going to do. (Year Nine Writing Test [session 2 of 2], p. 1, L 13-15)

TT: Right, coming round is a booklet – I need you quiet, please! As of now you are under test conditions, that means you do not speak, it means you listen to everything I have to say. Coming round is a booklet.

S: What do I need that for!

TT: For the moment, do not open it. (Year Seven Reading Test, p. 2, L 11-16)

At the same time as clarifying and communicating test requirements and expectations to students, TT focused students' attention to established classroom rules of behaviour and to those behavioural expectations codified through test instructions. T2 too noted that her experience with testing had taught her to attend to behaviour and classroom management in a preventative sense with her attention to "Classroom organisation (seating etc.)" (T2, Questionnaire, Q4). As such, teacher-student interactions through the test situation reflected patterns of behaviour that were both necessary for the administration and completion of tests, and consistent with established classroom practices and the test implementation strategies.

5.2.1.2 Affirmation: "do what you're told".

The data demonstrate that teacher-student action and interaction regarding instructions, rules and expectations during test sessions often went beyond patterns of clarification. Notable across the data within this pattern is actions and interactions by which rules of testing were affirmed. Within the overarching pattern of teacher-student action and interaction, TT's reiteration of rules governing test sessions and classroom behavioural norms was a frequently-noted teacher response to student acts of noncompliance. TT and his students discussed set rules and requirements at various times through test sessions, not only at the introduction of sessions. The selective emphasis and focus on rules demonstrated that participants considered those roles necessary for the administration of test sessions and in order to ensure the veracity,

validity and usefulness of test results. These included appropriate behaviour, quiet, and compliance with tasks.

The first element of this pattern pertains to affirmation of test rules at the outset and during test sessions. Within this pattern, TT articulated directly to students his concerns and requirements of their behaviour when introducing and administering the tests. The following extracts are examples.

- TT: [7MS1], please don't do that. Now, I'm just a little concerned about the level of calmness that's happening this morning. You're probably in a situation where you need to be rather careful that you don't end up losing some lunchtime to get this test done then. The last thing you want is for that to happen. So that means you need to be really calm and sensible about the way you behave right now. (Year Seven Spelling Test, p. 2, L 25-29)
- 8.52 am – 8.54 am instructions. Very clear. Students not listening, or still talking, are spoken to strongly. (Year Nine Reading Test, Field Notes, p. 12)

These extracts show that TT interacted with his students at both an individual and a whole class level regarding their behaviour and compliance with rules during test sessions. These examples show that TT provided qualified reasoning for students to comply with behavioural expectations as well as an unqualified request for compliant behaviour. TT opened test sessions by clarifying and reading to students the rules that would govern the task, then returned to such rules when students chose not to comply.

One key behavioural test rule noted across teacher-student test session interactions was that of silent student work. Throughout the testing, TT concentrated on students' listening skills and noise levels. The extracts below

are among many examples of this focus on, and reiteration of, the behavioural expectation of students being quiet during testing sessions.

TT: That means I don't want to hear another sound – as of now you are under exam conditions, which means that you do not speak –

S: ...

S: ...

TT: You do not speak, you do not spend any time worrying about what other people are doing, you are worried only about what you are doing. (Year Nine Writing Test [session 1 of 2], p. 1, L 10-15)

TT: Okay you need to listen please. That means from now you are silent – [9MS3], can you come and sit here please? Too many of you at that desk. Thank you, [9MS3]. (Year Nine Reading Test, p. 3, L 15-17)

TT: I'll tell you in a moment. Uh, you've got a choice here folks, this test takes more than an hour. If you end up in lunchtime, then you're gonna do it in lunchtime. If you don't listen now, I am not gonna start until you're prepared to be sensible. (Year Seven Writing Test, p. 4, L 9-11)

The data demonstrate that TT's stressing of student silence as a requirement for such situations as standardised literacy test sessions indicates his opinion that students' test experiences and behaviours influenced students' test performance. During test sessions, student-student interaction created such noise that TT responded by reiterating that the testing tasks were individual activities. Minimising student noise was also related in TT's emphasis on rules and expectations, to monitoring students' physical location in the classroom and to willingness to complete tasks. This latter point was evident through TT's reference to students' "choice" between being quiet, focused and "sensible", and the alternative of sacrificing non-class time to finishing the task; further, TT used time reminders as a behavioural corrector.

A further data pattern across this theme is that of reiteration of rules to students. Within this pattern, TT emphasised the importance he and the school overall placed on the upholding of rules in particular circumstances. Again, the extracts below are among the many examples of this pattern from through the data.

- TT: No! [7MS4], you need to sit, please. ...
- S: ... computer ...
- [Teacher calling attendance]
- [Student screaming]
- TT: [7MS4] – [7MS4], go and sit yourself over on the decisions desk please, over there.
- S: ...
- TT: No, over there, thanks. (Year Seven Writing Test, p. 2, L 13-20)
- TT: Leave them please, leave them.
- S: ...
- TT: Leave them please!
- S: ...
- TT: Because he did what he was told! That's the difference! (Year Nine Reading Test, p. 1, L 11-15)
- TT: That means that if you're finished, and you start making a noise, then you're being completely unfair to those who have not finished, and that's completely unreasonable. I can tell you that [P] is very serious about these tests being done properly, and I would think that anybody who chooses to disrupt a test like this would be choosing to get themselves into some pretty deep strife. (Year Nine Writing Test [session 1 of 2], p. 2, L 8-14)
- S: R-E-T-R-
- TT: [9MS12]! Isn't fair. (Year Nine Spelling Test, p. 11, L 10-11))

These interaction examples indicate that TT and his students drew on their shared history of interactivity in order to develop a working relationship. TT-student interaction concerning rule (non)compliance shows that TT chose to

provide ample warnings and reminders rather than immediately escalating conflict. Teacher-student interactions regarding rules and expectations only selectively escalated from reiteration of rules and requests for compliance into student relocation and threats of involvement with senior staff, through these examples, show that even during such interchanges TT worked to maintain positive relations with noncompliant students.

The above extracts, particularly the final two, further demonstrate two approaches TT chose to adopt in affirming test rules to students: appeals to authority and to students' concept of fairness. Through the former, P and other senior school staff represented such authority at the school site. The data demonstrate that TT was not averse to threatening students with P's involvement, should students continue to ignore his directions and test rules. Through his appeal to P's authority, TT emphasised the importance school staff placed on student behavioural and rule compliance during test sessions, and teacher-student interactions are shown to again adopt the use of student choice in behaviour during tasks. TT's reference to the idea of fairness to peers meant TT stressed that compliance with rules was 'fair' and 'reasonable', and that noise during test sessions would be neither fair nor reasonable for those still working.

5.2.2 Responding to student questions and concerns of testing.

Within the pattern of teacher-student action and interaction regarding student concerns and questions about tests, two pattern groups are identifiable. The first is of participant responses of negotiation and flexibility (5.2.2.1), by which

participants justified and engaged in breaking set rules when interacting with particular students. The second group of data pattern provides examples of academic and time management concerns, which were two key sources of student questions and concerns explored in teacher-student interactions (5.2.2.2).

5.2.2.1 *Negotiation and flexibility.*

5.2.2.1.1 *Justifying and providing assistance:* *"I don't see that as unreasonable".*

The level of assistance permitted for students during standardised literacy testing sessions was clearly problematic for the teacher participants. The first key pattern in this regard pertains to participants' references to student inexperience with testing activities, which they linked to a need for providing sympathy and, at times, assistance. The following are examples from each of the three teacher participants regarding this pattern.

- TT: So most of these students have never done a formal test like this before.
- R: Not even two years ago in year seven?
- TT: Oh, apart from that, apart from that. (Year Nine Writing Test [session 1 of 2], p. 4, L 12-14)
- T2: ...we've never had really exams or anything like that here, so this was the only sort of test, the only time when you really have to learn about skills of being quiet and um techniques of how to do it (T2, Int. 1, p. 2, L 20-23)
- T1: I mean I don't have any proof or anything like that, it just appeared that way to me. Whereas the older they got, the more stressful it appears to be. To the fact that some grade 9s basically refused to do it. Coaxed most of them into it, but some of them I couldn't get into it. When they knew it was the last day to do it, they didn't come. Things like that. (T1, Int. 1, p. 2, L 26-31)

These particular extracts indicated that the participants shared the opinion that student anxiety was a direct result of standardised literacy testing, caused either by inexperience (as suggested by TT and T2) or by test-induced pressure caused by increasing exposure (suggested by T1). Student anxiety and lack of skill in such a “formal test” situation appeared to contribute to the participants’ conflicted opinions regarding the amount of help they were able to provide their students.

Student test experience was notable through students’ questionnaire responses also, with the observed Year Nine students overwhelmingly responding in the negative when asked if they or their teachers follow up on the tests (17 negative and no positive responses), compared with the Year Seven students who responded largely in the positive (15 positive to two negative responses). Also notable was the difference in overall responses when asked if they or their teachers take the tests seriously. The Year Seven students surveyed were overwhelmingly positive about their opinions on the tests (11 positive and three negative responses) and about their teachers’ predicted opinions on the tests (12 positive to two negative responses). The Year Nine students surveyed were more mixed in their opinions, with positive responses regarding their own considerations of the tests (16 positive and one negative response) but less consensus regarding their predictions of their teachers’ opinions (nine positive, two negative and five unsure responses) (see Appendix C).

A similar pattern across the data regarding students’ testing confidence pertained to teachers’ responses. TT, T1 and T2 appeared to differently

problematise and respond to the experience and difficulty of aspects or issues of testing. The extracts below provide examples of this pattern.

TT: So, it's just necessary, it's expedient, to provide more than the basic instructions that are provided in the test itself. And if you didn't do that, then you would just be setting yourself up for disaster. In doing that, you try very hard not to influence the result that the individual gets, in that your explanations have to be explaining the question, rather than explaining the answer. That's often quite difficult to do.

R: Really difficult, given some of the questions they were asking.

TT: Yeah. So it's often quite difficult to do, but you've gotta do your best to do that, in order to make the test relatively valid. (TT, Int. 3, p. 5, L 17-25)

T1: Yes, it seemed like that's the first time they'd had to do one of these, they were used to calling out in class 'what's the answer? What's the answer?' They want you to tell them the answer, and when you tell them you couldn't and they – it was quite distressing for some of them. And it was really quite difficult. (T1, Int. 1, p. 1, L 4-7)

T2: To be able to do the tests properly. Because I do want to make it possible that they can do the best that they can do (T2, Int. 1, p. 2, L 1-2)

The role inflexibility brought on by being test administrator and overseer meant for T1 an inability to move from being the administrator to being a helper. In comparison, TT and T2 were more willing to overstep this role restriction, and justified their preparedness to disregard the assistance rule with arguments about test validity and helping students, explaining that providing limited assistance ensured that students could understand and respond to test tasks. Those two shared a concern here for ensuring the accuracy and reliability of test results, and, in this regard, sought to avoid providing the kind or amount of assistance that might skew results.

Participants identified that issues within test content and processes served to justify their assistance of students during sessions. This pattern was particularly evident across TT and T2's expressions, and extracts below are examples of their discussion of this issue.

TT: But often, the language of the testing is the language of the people who write the test, rather than the language of the individuals being tested, and it's a bit like IQ tests; they're fraught, because of the issue of the way the questions are asked. Someone's IQ might read quite low, simply because they didn't understand the question. (TT, Int. 3, p. 5, L 25-29)

TT: I think there is a real need for several versions of the same test, yeah, because for some individuals who almost can't read, you are guaranteeing their failure by presenting them with a test which is way too hard. If you presented them with a test that was maybe the test that the kids in grade three do – but, perhaps, the topics might be changed, but the quality of the questions, the difficulty of the questions, was about grade two or three standard, then there would be quite a few people who would do a fair bit more in the tests, and would reveal more about themselves in the process. (TT, Int. 3, p. 10, L 6-13)

T2: ... one of them was really not well organised, I think it was that one, that if you just did what it says, the kids would have been quite confused, I think. So I don't follow every single thing to the letter. (T2, Int. 1, p. 1, L 25-27)

TT and T2 were both aware that providing assistance to students during test sessions was discouraged in testing guidelines, and both wanted to avoid influencing students' results, providing assistance for students only at the minimal levels required to achieve student understanding of test questions and tasks.

TT noted that student frustration with test design and content was particularly evident; "Many dislike the language and style – particularly

spelling” (TT, Questionnaire, Q6a). Related to this broad pattern is that of test structure, content and design, which the following extracts identify as a further cause of frustration for teachers and students.

- TT: This would be the worst test to administer, because the words weren't numbered. And they're assuming that the kids can keep up with the passage that's being read, rather than telling them the number of word to fill in. Made it probably twice as difficult to do this test as it would have been last year. (Year Nine Spelling Test, p. 13, L 28-31)
- T1: I read through them first, then I read through them to the kids. And then we picked out the bits, and then we – I did those bits again. I would imagine – I mean, I'm not very good at doing the same thing over and over again. That's not my forte. I find that very difficult to do things identically so I would probably do things differently almost every time, to a degree. But I would think that it would be something along those lines, if I'm doing that again. (T1, Int. 1, p. 9, L 9-23)
- Hints with techniques (e.g. reading questions before reading the passage). (T2, Questionnaire, Q4).

Allaying student confusion in turn meant that students could understand tasks and, perhaps, complete the tests as they would other assessment tasks, producing responses that accurately and genuinely reflected their capabilities and skills. This meant providing further instruction than the test guidelines required. TT's discussion of his desire to allay student confusion had clear consequences for following instructions. He referred to the alternative, of not providing additional and unsanctioned instruction, with an allusion to clearly negative classroom management repercussions.

TT and T2's reiteration of the difficulty of minimising influence upon testing and maximising test validity suggested pressure on participants; they

were expected to ensure students completed the without enabling inflated results, though difficulties in achieving this balance were evident.

5.2.2.1.2 *Teacher-student rule negotiation and flexibility:
“negotiating with individuals” / “you learn not to raise your voice”.*

Negotiation over rules and behaviours was a major way in which TT and his students sought more constructive and mutually preferable interactions across the test situation. Such negotiation appeared to be a manifestation or result of his already-noted perception of what undertaking these tests meant, in practice, for his students. Participants’ justification of their classroom management and student interactions with reference to the student groups’ readiness to engage in conflict has already been noted.

The first pattern within teacher-student interaction over rule negotiations was that of students expressing their unwillingness or inability to continue to work on their tests, and TT’s suggestion of an alternative activity, such as moving to a computer area so they did not distract other students in the main classroom area. The following extracts are among the numerous examples of TT’s negotiation with students to move to a designated computer area supervised by an Assistant Principal, here referred to as AP.

- TT: – You need to be able to find something to do that’s going to be able to keep you out of trouble –
- S: Computer.
- TT: – while others are working. (Year Nine Writing Test [session 2 of 2], p. 1, L 28-31)
- TT: [to another student – 7MS1?] When you finish you can go to [AP].
- S: M-hm.
- TT: Yeah, but, you’ve gotta do your very best.

[extended period of relative quiet; some whispering]

TT: [to R] We've bribed them with being able to play computer games so they do the right thing. (Year Seven Writing Test, p. 13, L 23-28)

The above examples show that TT chose during test sessions to focus his attention on the students in the classroom instead of having to divide his attention between two or more areas. TT and his students were clear that acceptable behaviours could be negotiated through discussing alternative activities (such as using a computer) and given student assurances of nondisruption.

A second pattern of rule negotiation and flexibility, related to that presented above, showed that TT also negotiated rules and behaviours as a form of reward. Students who worked hard for what TT considered to be an acceptable amount of time or produced a particular amount of work were allowed to negotiate or be granted a break, as the extracts below example.

TT: How are you going? You're doing a good job. Got plenty of time. Have a break if you need to. [to 9MS1] (Year Nine Reading Test, p. 7, L 15-16)

TT: Yep. [7MS5], [7MS1], listen, go on with ... you're doing a good job. Okay? Right, you've done a good job, you can go off to [AP] if you like.

S: ...

TT: Yeah you've both had a go, you can both ...

S: Can [7MS1] ... if [AP] says he can?

TT: Um, when [7MS1] has done a bit more.

S: How much more do I gotta do!

TT: ... you've gotta finish ten more lines.

S: Ten more lines?!

TT: Yep.

S: [counting one to ten out loud] (Year Seven Writing Test, p. 12, L 1-11)

TT and students appeared experienced and comfortable with this kind of reward system; the data showed the students to be practised in such negotiations with their teacher. As already noted, TT clarified on a number of occasions to the researcher that interaction regarding negotiation of rules and acceptable behaviours were necessary with these students when engaging in lesson times, rather than strictly-imposed rule reinforcement. TT justified his approach of ignoring student misbehaviour through reference to the demands and culture of the particular student group.

TT repeatedly manifested his belief in negotiating with students to avoid conflict. During the test situation, he and his students were often engaged in discussions regarding how, when and for what reasons students would be punished. The following extracts example that rules were clearly stated for students, so they were aware of what was negotiable, and following such clarification, negotiation over rules and TT's flexibility could be seen.

TT: ... I asked you not to speak, go over to the decisions desk.

S: No! ...

TT: Go and work over at the decisions desk.

S: I'll be quiet.

TT: You will! (Year Seven Writing Test, p. 10, L 5-9)

TT: ... quiet, don't disturb anyone else. Right, in the next minute or so you need to tidy up what you are doing. If you need more time, then you can have another twenty minutes the next time I see you. If that's the case, if you have not finished, when you give me back the booklet, can you leave the magazine inside your booklet, so that I know you're one of the people who needs extra time. (Year Seven Reading Test, p. 11, L 4-8)

TT: Get to work please. When you've finished, booklet comes to me.

S: I'm not in the right kind of mood to write now ... (Year Nine Writing Test [session 2 of 2], p. 2, L 10-11)

The “decisions desk”, mentioned above and in previous pages, was a single desk and chair, placed near the classroom door and away from the main classroom area. Relocation to the decisions desk represented a behaviour management strategy, to which the students were accustomed. TT and students demonstrated their familiarity with negotiating various aspects of the task, such as rules for completing the task and acceptable student behaviour. TT and his students were comfortable discussing how, why and for what infractions the students could and would be punished.

Before escalating his behaviour management steps, TT clarified the specified rules for the test situation, so students understood the parameters, what they could and could not attempt to negotiate. The data provide examples of student noncompliance in the form of minor infractions, such as quiet talking or expressing a dislike of the tests, which TT chose to ignore, or not escalate into conflict. The extracts above serve as examples of teacher-student interactions in which students made clear to TT their decision not to comply with stated rules or expectations, to which TT gave a considered response.

A further pattern of test rule flexibility and negotiation is evident across TT’s decisions to ignore acts of student rudeness and noncompliance through test sessions. As noted, during individual teacher-student interactions TT showed himself willing to ignore student noncompliance and inappropriate language. The following extracts are examples of TT’s willingness to ignore student rudeness directed at him and student refusal to engage with tasks.

TT: Some of the words have been left out.

S: Der.

TT: First I'm going to read the whole thing to you, including the missing words, so the copy I have has all the words. (Year Seven Spelling Test, p. 4, L 14-18)

S: There's a mistake in this. [pointing out a spelling error in booklet]

TT: Don't worry about that.

S: Give us a look.

[students talking about spelling error]

TT: [9MS8], shh! (Year Nine Reading Test, p. 4, L 12-16)

- [9MS5] and the new boy are whispering; teacher says he will move [9MS5]. New boy checking his mobile, which is in his pocket. [9MS5] had said to himself that he needed to put his on 'vibrate'. (Shouldn't phones be banned from the classroom, or at least turned off during the test?)
- Girl at right back closest to the aisle has earbuds in: don't know if she is listening to music. (Year Nine Writing Test [session 1 of 2], Field Notes, p. 8)

These extracts show that TT selectively disciplined students. TT demonstrated his occasional considered willingness to discipline rather than appease or ignore noncompliant and rude students, with the above extracts highlighting a similar pattern of such interaction. In the instances exemplified above, the students deliberately ignored TT's clear direction and continued with their conversations, and TT's apparently minor reactions suggest that he was not at all concerned with the student noncompliance, and that he chose not to escalate the incident but to continue with the test task.

During test sessions, some students' test avoidance was ignored or unnoticed by TT who was otherwise occupied by having to attend to controlling disturbances with potential for dramatic escalation. The data suggest TT was experienced with these students and considered his approach to rules and noncompliance in light of this experience. At times, TT's ignoring of

students' noncompliance with his directions and instructions happened accidentally, without TT's realisation of continued student noncompliance. This data pattern suggests that TT did, however, work to maintain control in the classroom where possible, for example through sympathetic interactions with students, demanding that TT react selectively to student disturbances.

One Year Nine student in particular who had struggled throughout the testing tasks seemed to have difficulties understanding and following TT's directions. That student, given the pseudonym '9MS1', was frequently engaged in interaction with TT regarding rules and expectations; the example of this student, with quotes from TT's audio recorder and the field notes, demonstrates 9MS1's manifestations of anger and frustration, and TT's considered, experienced and sympathetic response.

- "I am not writing this. [You/they] can go and get fucked." (Year Nine Writing Test [session 1 of 2], Field Notes, p. 7)
 - Had to tell [9MS1] he was doing the wrong section: "Bullshit. I'm not very happy about that." (Year Nine Writing Test [session 1 of 2], Field Notes, p. 7)
- TT: ... And as he said to me: 'you're not in my situation, put yourself in my shoes', or whatever he said, and I agree with him. He sat there for thirty-five, forty minutes, knowing that he couldn't do it, knowing that all the other kids knew he couldn't do it. He wrote five words. So you feel very sorry for him, but he's put in the situation because of the demands of the tests. (Year Nine Writing Test [session 1 of 2], p. 7, L 20-24)
- [9MS1] is functionally illiterate: very demeaning. Everyone knows he can't do the task; no wonder he's disruptive to the other students. (Year Nine Writing Test [session 1 of 2], Field Notes, p. 9)

Again indicating TT's experience with the students and the nature of the student-teacher relationship, these examples above show that TT chose not to react to the performance, even when confrontational language was directed at him. TT identified this student as struggling with the requirements of the task, and made a clear decision about the method he would employ to handle the situation. 9MS1's reasons for noncompliance and disruptive behaviour in the testing situation were readily identifiable to the researcher when observing the session, and the following excerpt from the field notes echoes TT's similar comments. With behaviour directed at TT, the student's test noncompliance above that appeared to be directed at TT indicated his disinclination to complete the testing tasks. TT's method of response communicated his understanding and possible empathy about how such students felt about testing. The situation was characterised by calm, empathetic response to the student, communicating TT's interpretation of that student's situation.

TT demonstrated his belief in the importance of carefully evaluating incidents to determine an appropriate response, for example acknowledgement or escalation. The data indicate that this approach of selectively ignoring student test noncompliance was at times effective in minimising disturbances and encouraging conscientious students to continue their work uninterrupted. The data indicate that TT's reactions to student test negativity were influenced by his opinions of and previous experiences with the students, which enabled him to make decisions regarding which acts of student test negativity or noncompliance to more immediately address and how best to do so. The data

highlight that TT presented himself as flexible, in order to minimise conflict and maximise completed work throughout the test situation.

5.2.2.2 *Student concerns:*
“I can’t read that” / “What if we can’t do it in two lessons?”.

Teacher and student actions and interactions throughout the test sessions reflected their shared concerns about the students’ abilities within the testing context. The first pattern within this theme is that of students vocalising to TT their concerns about the tests. Such interactions frequently took the form of teacher-student discussions regarding the set tasks. Students particularly questioned TT about definitions of terms within test tasks, how they should respond to the tasks, and time allocations for tasks, and clearly stated to the teacher that they believed themselves incapable of completing tasks and queried repercussions should they be unable to complete.

Within the actual test situation, the teacher-student interactions focused on student test questions and concerns but also reflected time concerns, particularly on the part of students; the examples of this pattern demonstrate that TT used time reminders to both re-focus students to the task and to provide encouragement, and to selectively (re)orient the students’ focus during test sessions. The extracts below are among the large number of examples of this pattern.

- S: What if we can’t do it in two lessons?
- TT: If you do – huh? You will have to do it in two lessons because
 you only get a certain amount of time to do it.
- S: But what if ... (Year Nine Writing Test [session 1 of 2], p. 2, L 2-5)
- S: What is the text?

- TT: The text is the writing.
S: Okay. (Year Seven Reading Test, p. 9, L 1-3)
- TT: There is of course no penalty for getting the words wrong.
Remember this test does not count towards the end of the year, so
it will not matter, provided you have a go. (Year Nine Spelling
Test, p. 4, L 7-9)
- TT: This, you will be pleased to know, is the last of the tests. [some
'yays' from students] This one is – well it's the simplest of the
tests in that all you've got to do is do some spelling. So you can
either do it or you can't. (Year Nine Spelling Test, p. 2, L 14-17)
- TT: The words are varying in difficulty: some of them you'll have no
trouble with; some of them are hard for almost anybody. (Year
Nine Spelling Test, p. 3, L 29-30)
- S: Does it have to be finished by this lesson?
TT: Before lunch, yeah. Got plenty of time.
S: So we can take all the rest of this lesson?
TT: If you need to. (Year Seven Writing Test, p. 16, L 19-22)

Much teacher-student interaction appeared to be directed towards ensuring students were able to engage with and complete their work. Examples such as those above show TT was at times able to provide students with sufficient information without subverting in a major way test rules. The examples above further demonstrate that the observed students continuously communicated their concerns to the teacher throughout test sessions, indicating students' comfort in expressing their academic and time management concerns, vocalising distress at test tasks to TT during the test sessions. TT's information did not always calm his students, however, who continued to ask about time and consequences if tests remained unfinished at the end of sessions.

This pattern demonstrates TT's awareness of his students' academic concerns, and his attempts at times to allay them. TT attempted variously to reassure students that there would be no repercussions for misspellings, but nonetheless emphasised their concerns by himself choosing to orient his introduction around the tests' difficulty level. The above examples of teacher-student interaction further present that students were often so concerned about the standard of the work they were expected to produce that the teacher minimised the work expectation.

When introducing the test sessions, TT's references to the academic difficulty of the tests in contrasting ways often adopted overtly positive language. TT's discussion of time, specifically time limitations, assigned for the tests was both a reassurance for concerned students and a control for student behaviour. TT's words to students about their academic ability to complete the tests reassured students that they would not have to plan, produce and perfect a response, and that their ability to complete the task would not depend on an outlay of effort but rather would depend on skills they did or did not already possess. TT extended the idea of tests as merely gauging already-developed abilities rather than challenging the students. However, TT's focus on student difficulty in completing test tasks simultaneously affirmed to students that they might find the tests difficult, although students were encouraged to reconcile themselves to the tasks. TT did not claim that the spelling tests would be undemanding, but prepared students for the challenge prior to beginning the tests.

The second pattern identified across this theme highlights that, in spite of student and teacher awareness of academic concerns about test performance, TT was often either unable, or chose not, to respond to students' questions or requests for assistance. The following extracts are examples of TT's response to such occasions, which usually either saw him imply that the answer would be forthcoming, or that he was unable to provide the requested assistance and that students should do their best.

- S: What do we do about the characters?
TT: Just give us a second and we'll get to that. (Year Nine Writing Test [session 1 of 2], p. 3, L 11-12)
- TT: "Their focus". You've got to think about which "their".
S: Can we just write all three?
TT: Can't write all three.
S: How many are there?!
TT: Shh-shh-shh. Shh, [9MS3]! (Year Nine Spelling Test, p. 8, L 1-5)

These above extracts demonstrate that students' academic test concerns ranged from queries that could be answered or minimised by a simple reiteration of rules to more fundamental student anxiety about their academic capacity. The above extracts further demonstrate that students' anxiety about their academic ability to complete testing tasks sometimes disturbed the test sessions and resulted in TT's reiteration of behavioural expectations.

The third identifiable pattern within test concerns a further use of time reminders during test sessions. As well as providing assurances and answering queries about time allocations, TT also used references to time as part of a behaviour management strategy. By employing a clear reminder of remaining

time, TT could reiterate his behavioural expectations. The following is one clear example of this approach.

TT: We're running out of time, folks. Means I do want silence, and I mean it. (Year Seven Reading Test, p. 2, L 25-26)

Students' time concerns were at times exacerbated and legitimated by TT using time frames as a threat, seen in the above extract. Yet, at other times, TT attempted to minimise these same student time frame concerns. Context appeared to dictate the nature of time-related interactions. The data show that discussion of rules and time requirements relating to tasks occurred when students sought clarification about these from TT, and in addition, TT used reminders as correctives to student noncompliance in this pattern.

The fourth pattern within this theme demonstrates that, at times through test sessions, teacher-student interactions regarding academics and time were not initiated by teacher or student concern. In some of the teacher-student interactions, students sought merely a clarification about time (time allocations for tests, and time remaining for particular sessions), rather than general encouragement about their ability to complete within such time frames. The example below demonstrates this pattern.

S: We're doing it now?
TT: Oh we won't do it right now, we'll have a break for half an hour.
S: How long will it go for?
TT: Oh, it'll take – fifteen minutes?
S: Okay. (Year Nine Reading Test, p. 8, L 22-26)

Examples such as that above demonstrate that TT's student group had differing levels of confidence when approaching test tasks. Indeed, the data demonstrate that not all students were overly concerned with time issues. This is another

element TT needed to consider in selecting his test administration technique, and explains TT's evident consideration of his experience with these students in making decisions of whose queries and behaviour to focus on and which aspects of test tasks to emphasise and when.

5.2.3 Addressing student negativity towards testing.

At times, both TT and his students expressed negativity towards testing. Students expressed negativity towards the tasks they were expected to complete, the rules they were expected to comply with, and their lack of choice in whether or not to participate in testing. Students both verbally and behaviourally expressed anger at the testing situation, through their interactions with each other and with TT. In response to such demonstrations of frustration and noncompliance, TT also demonstrated his text-negativity, which he couched in humourous and encouraging interactions with students. This subsection presents and explains examples of such observed actions and interactions.

5.2.3.1 Humour: "Throwing pencils at you!"

Teacher-student interaction in the standardised literacy test situation often included jokes, humour and a relaxed approach. Humour employed by TT actioned and reinforced his viewpoints regarding the importance of avoiding conflict. Humour appeared to be welcomed by the students, who often returned jokes and engaged in friendly, relaxed interactions with their teacher and each other. The following extracts present examples of such humourous interchanges between TT and his students.

- S: Hello!
- TT: Yes, it's listening to everything you say! (Year Seven Reading Test, p. 2, L 4-5)
- TT: Does anybody have the wrong answer sheet or test booklet?
That's the first question.
- S: I've got the wrong answer sheet.
- TT: Does it have your name on it, [9MS5]?
- S: Nuk.
- TT: That's where it says '[9MS5] –'.
[students laughing] (Year Nine Reading Test, p. 3, L 26-32)
- [TT accidentally drops box of pencils]
- TT: Who-ho-ho, that was good! Eight pencils! Throwing pencils at you! (Year Nine Spelling Test, p. 3, L 9-10)
- TT: Oh lucky her, a test on her birthday! What sort of a treat is that!
(Year Nine Reading Test, p. 10, L 21)

The data show TT humour when interacting with students at all stages of the test sessions, and that he employed the humour to maintain positivity in test sessions and to encourage test completion. TT engaged in joking with students before and following testing sessions. He used humour to explain test rules and related classroom procedures, and to reinforce desired student behaviour . TT's humour when administering the tests communicated the message that the testing sessions would only take minimal time and need not be overly unpleasant. The humour was often in the form of jokes about the tests, implying that he considered the tests uninteresting. TT and his students often comfortably joked together, indicating that humour was part of their classroom culture and relationships, not limited to test situations. The use of humour seemed to appeal to students as much as it did to TT, facilitating a positive classroom environment.

5.2.3.2 *Encouragement: “really pleased to hear that”.*

TT encouraged students in a variety of ways. Encouragement took the form of offered breaks to students he perceived as working conscientiously, assistance to students struggling on the tests, suggestions to students that they had underestimated their ability, positive acknowledgements of whole class efforts, and praise for student efforts following test sessions.

The overall pattern of student encouragement was couched within participants' expressions of students' desire for academic success. The two most overt statements regarding this pattern are given below.

TT: There is a very small group of students who will do nothing because they just choose to do nothing. Most kids, most of the time, if they can do the task, will make a reasonable attempt. (TT, Int. 3, p. 9, L 20-22)

T2: I think most of the kids I teach actually do quite well if they can, but they really seem, even the laid back grade nine boys, they put their heads down and they get really worried if they're not sure if they're doing it right, so they do want to succeed (T2, p. 10, L 28-31)

This opinion expressed by both TT and T2, that all students wish to academically succeed and only need encouragement to do so, contextualises and explains the general pattern across observed teacher-student interactions of encouragement of students.

Teacher encouragement of students during test sessions, in addition to TT's use of humour in student interactions, saw TT rewarding student efforts with breaks from tasks and positive words. The extracts below provide examples.

- TT: Very funny. That is magnificent, [7FS9]! You deserve a break, [7FS2]. Do something quiet, [7MS1]. [7MS15], if you want to use a computer you can, long as you're quiet. Same applies to you. (Year Seven Writing Test, p. 19, L 10-12)
- TT: Oh I think you can do more ... [7MS4].
- S: No! Bad, bad, bad, bad, bad.
- TT: Doesn't matter about spelling. [pause – then under breath] Pulling teeth. (Year Seven Writing Test, p. 11, L 17-19)
- A2: I didn't realise we had so much time. [whispering] You could probably do another ten lines.
- S: ...
- TT: Right. You've done it. You can go.
- A2: Well done. (Year Seven Writing Test, p. 15, L 1-5)
- TT: If you need an eraser because you think you've made a mistake, you need to put your hand up and I will get that to you. Secondly, the answers to every question *are* in the thing that you have read. So if you read it carefully you *will* find the answer. (Year Nine Reading Test, p. 5, L 33 – p. 6, L 2)

These positive forms of teacher-student test interaction as the above examples were particularly employed with students who appeared to be struggling to complete the testing tasks. During teacher-student discussions, TT overtly helped students feel more confident in their abilities and these examples suggest that TT empathised with his students. TT encouraged students to feel capable of completing their work, making deals with some students to ensure their compliance. This negotiation for increased student productivity was a further expression of sympathy and encouragement by TT. In addition to encouraging statements to struggling or unmotivated students, TT established his expectation at the beginning of test sessions that students would work to the best of their ability.

The third pattern across this theme is of encouragement of students for their performance during test sessions. Within this pattern is identifiable both praise for individual students' efforts at the conclusion of each session, and follow-up praise at the beginning of sessions. The following extracts are among the examples of this pattern.

TT: Yesterday you were *excellent* and you allowed the people who had not finished to do their job really well. It's really important that you continue to show that sort of maturity and consideration for each other, so that everybody can do their best. (Year Nine Reading Test, p. 5, L 21-24)

TT: Yes, you might get a certificate for putting in a big effort, I reckon, [9MS1]. (Year Nine Spelling Test, p. 14, L 10)

TT: [7FS2]. You have done such a good job today. Should be really proud of yourself. (Year Seven Writing Test, p. 24, L 3-4)

As noted through the above examples, TT followed up on test sessions by praising students for their focus and performance at the beginning of the following test session. This positive introduction to the test sessions appeared to encourage students. These extracts exemplify teacher-student interaction patterns already noted across the data. They communicate that TT valued behavioural compliance and its attendant affect upon student test performance and results. As such, the above instances of teacher praise for student behaviour also contain an inherent rebuke for those students who might not have engaged in the expected behaviours.

Testing tasks are valued by TT in the above extracts, as sufficiently important in themselves to generate positive recognition for appropriate and compliant student action. Collecting test booklets from students at the

conclusion of the test sessions was often used by TT as an opportunity to praise student efforts. Such praise as exemplified above appeared to facilitate a positive conclusion to the test sessions for both TT and his students. Further, the first extract above, demonstrating follow-up praise of students, also extend the ‘fairness’ and ‘reasonable’ pattern already noted, to include a valuing of “maturity and consideration”, of students acting and interacting in such a way as to not infringe upon other students still working.

5.2.3.3 *Teacher test-negativity: “gonna have to bear with it”.*

TT communicated his negative opinions of standardised literacy tests and testing to his students at the same time maintaining encouragement to the students. This broad pattern across teacher-student interactions during and regarding test sessions included TT statements that amount to his negative opinion of test tasks, content and requirements.

Teacher-student test interactions were responsive to student test-negativity and sometimes clearly derogatory attitudes about standardised literacy tests were communicated. TT at times responded to student test-negativity with a matching negativity. He expressed his own negative views about testing in a number of ways, but appeared to incorporate some encouragement of his students within these statements. TT and his students often discussed the tests and, during those interactions, as has already been noted, TT expressed negativity about the enjoyability of the tests, the administration, and the final usefulness of standardised literacy testing. Despite making essentially negative comments regarding standardised literacy

tests and testing, TT nevertheless incorporated an element of encouragement for his students.

TT: An appropriate word to begin with: “boring”. (Year Nine Spelling Test, p. 13, L 6)

TT: Yeah mouths shut please, I know this is difficult, but you’ve gotta do your best. (Year Nine Spelling Test, p. 12, L 3)

TT: Now, I am aware that for some of you, this is going to be a task that you don’t enjoy, but everybody in Tasmania has to do this, you’re not the only people. (Year Nine Writing Test [session 1 of 2], p. 1, L 15-17)

TT: I realise this is not the easiest way to run a test, but you’re gonna have to bear with it and do your best, we will only be another five or ten minutes. (Year Nine Spelling Test, p. 9, L 14-15)

TT was both negative about the tests and encouraging of students, communicating cynicism and complicity. TT’s comments implied that the students were expected to be struggling, whether academically or behaviourally, but that if they could remain focused, the tests could be completed. TT’s point about the tests being boring and difficult was furthered by his emphasis that the tests were an enforced requirement. Through his language, TT anticipated, incorporated and perpetuated the idea that the tests would not be enjoyable. TT did not choose for the students to be tested in this way, but was forced to administer the test just as all students in the State were forced to undergo the testing. TT synthesised his sympathy for students, criticism of test tasks, and shared desire for the test to end.

5.2.4 Summary of research question 2.

Research question 2 was addressed through an examination of the patterns that constitute themes 1, 3, 4 and 5. The various patterns of test session action and interaction were identified and considered, including patterns observed by the researcher during test sessions. The actions and interactions, apparent and claimed, appeared to enable and facilitate test session administration as well as served to address a range of teacher and student concerns. Most important in this respect were issues of student confidence in approaching test tasks and teachers views about how best to ensure that test responses and results would be fair, valid and valuable. As such, the data indicate an overlap between these research questions, with an at times conflicted importance placed on test validity, which itself indicates an albeit conflicted perception of test value, all strongly tied with the participants' predicted uses of the test data and awareness of issues in test administration.

These participants demonstrated an array of methods and skills used to achieve particular outcomes in the classroom during the test sessions, including humour and discipline as determined by the context. Patterns of action and interaction identified as characterising test sessions reflected participants' often-contradictory opinions of the tests, at once both critical and positive of elements within testing and the tests themselves. Despite considerable reservations, these teachers were committed to encouraging their students' improvement in test-taking skills and behaviours.

The data here indicate that participants made considered and context-specific responses to the test situation, in which they weighed up test

expectations against perceived student needs and classroom realities. Here the data too show participant frustration at a perceived lack of public or system-level appreciation of the complexities inherent in the teachers' roles in testing, both valuing and utilising testing but aware of the limitations of test administration and student reactions.

5.3 Results

As the data regarding research question 1 indicated, these participants held conflicted opinions as to whether the tests and testing overall represented a positive aspect of their teaching obligations. Standardised literacy testing mostly represented, although to different degrees, some useful information about students' literacy skills couched within broader test processes that often affected what these teachers could actually do with the knowledge.

Standardised literacy testing, too, represented opportunities for students to demonstrate skills without utilising technologies to hide weaknesses, and for students to develop confidence in performing under test conditions; testing also represented pressure to conform to imposed rules and practices that constricted some students' activities to succeed on tasks. Such conflict within testing was further apparent in the observed testing sessions, where inconsistencies regarding test rules and procedures were demonstrated in practice. Participants were observed attempting to maintain a consistent classroom environment, especially based on productive teacher-student relationships during the test sessions, often in disregard of official guidelines and instructions.

These patterns repeated themselves through the data pertinent to research question 2. Test session patterns of action and interaction highlighted significant tensions for teachers within testing. Such tensions were notable through TT's classroom-site response to the struggle between demands for testing uniformity and teacher desire for reflexive teaching. Within the context of these data, reflexivity in teaching and classroom test administration refers to test actors' capacity to utilise experience and autonomy in making decisions, acting and interacting. These data confirmed the participants' tendency to prioritise a working classroom environment and positive teacher-student relationships over compliance with test guidelines with which they differed and often disagreed.

In light of the coding analysis and through addressing the two research questions, the initial results are:

- Participants considered, and accounted for, the value and validity of standardised literacy testing in a variety of ways, adopting contrastingly positive and critical stances;
- Participants identified factors that influence test responses, and expressed concerns about their influences upon the validity and accuracy of students' test results, through action or inaction;
- Participants valued testing and test data as a validation of their teaching (and assessment) judgements;
- Participants expressed clear frustrations and concerns about uses made of testing and test data, including political point-scoring, teacher deprofessionalisation and school comparisons;

- Participants attempted to address student concerns during test sessions in such a way as affected their opinions on the accuracy of test results and effects of testing on students;
- Participants made considered, context-specific responses at all stages of the testing process, balancing test expectations against perceived student needs and classroom realities, and further identified context-specific suggestions of issues that if addressed could improve the tests' value and validity at the school site, such as time frame issues;
- Participants expressed a level of frustration at a perceived lack of public or system-level awareness of the complexities inherent in teachers' roles in testing.

This results chapter has clarified and explored the patterns within the five themes and in response to the two overarching research questions. This chapter has highlighted the disparities between what participants imagined as the public 'ideal' of standardised literacy testing and what they experienced as the school-level 'reality' of testing. This disparity between the testing ideal and classroom reality was suggested through the data as causing a degree of frustration for the participants, leading to both positive and negative conclusions about standardised literacy testing.

These participants experienced particular conflicts when applying uniform conditions in their individual classroom environments. The data demonstrate that these participants appreciate the value for test results of ensuring test conditions are fair and standardised, but that they also considered that denying some help to struggling students resulted in tests results nonreflective of students' literacy levels. As such, through the data can be

noted something of a conflict between teachers' desire to utilise, on the one hand, their preferences for professional autonomy and reflexivity in their classroom sites, and on the other, the expected or demanded uniformity of test administration. These discursive tensions are examined in the following chapter, Chapter Six: Discourse Analysis, and considered within the debate and practice dilemmas.

Chapter Six: Discourse Analysis

Introduction

This chapter presents the findings of the discourse analysis of the data.

Chiefly, these findings show that the teacher participants' perceptions of and approaches to standardised literacy testing were characterised by two dominant discourses and the discursive tensions apparent between them. The discourses were identified through the constructivist grounded theory coding analysis (Charmaz, 2001, 2005, 2006) and through the two research questions, 'How do teachers perceive and interpret the value and validity of standardised literacy testing?' and 'What patterns of action and interaction characterise standardised literacy testing?' The identified discourses were analysed through an approach shaped by the works of Carabine (2001) and Fairclough (2001).

Chapters four and five presented the detail and results of the constructivist grounded theory coding (Charmaz, 2001, 2005, 2006) stages of the analysis of the six participant interview transcripts and six testing session transcripts through the frame of the research questions. Those chapters concluded with the identification of the initial analysis results and the noting of the discursive tensions within these participants' perceptions of standardised literacy testing as an educational tool and practice, and their approaches to the tests' administration. These tensions are those between system-level demands for uniformity in standardised literacy testing, and (school) site teachers' desires for reflexivity. This system-site binary (Freebody & Wyatt-Smith, 2004; Williamson & Myhill, 2008; Wyatt-Smith, 2008) highlights the often-

competing standardised literacy testing paradigms within which such teachers operate.

This chapter defines and examines the dominant discourses, and the tensions between and resultant of them. In doing so, the chapter examines the strategies and techniques by which the discourses are manifested and empowered, the effects of the discourses, absences and silences within the discourses, examples of an acceptance and a rejection of the discourse, the purposes served by the discourse, and identifiable obstacles to the particular problem. These details were explained in Chapter Three.

The findings of the discourse analysis provide a richer understanding of the meanings that can be made from the research study and site. The discourse analysis approaches employed in this research involve the questioning of the ways that the practices either reinforce or challenge hegemonies and norms, and of the social implications of the discourses in the wider power contexts (Carabine, 2001; Fairclough, 2001; Rowan, 2001; Taylor, 2001a, 2001b). The employed analysis approaches invite examination of representations (who or what is represented or included, in which ways, by whom, for what purposes, and to what effect), and sites of contestation (particularly the ways that participants are positioned, by whom, for what purposes, and to what effect) (Phillips & Hardy, 2002). A combination of discourse analysis approaches enables more thorough critique of the data, the topic of research, the meanings, patterns, themes and discourses contained therein, and the broader issues of interest (Anderson, 2004; Charmaz, 2005, 2006; Charmaz & Mitchell, 2002; Dey, 1999, 2007; Haig, 1995; Whipp, 1998; Yin, 2006); “different perspectives

provide different forms of knowledge about a phenomenon so that, together, they produce a broader understanding” (Phillips & Jørgensen, 2002, p. 4).

This discourse analysis considers the social implications of the discourses in context (Carabine, 2001; Rowan, 2001; Taylor, 2001a, 2001b). This discursive stage of the analysis critically contextualises and examines the discourses and discursive tensions identified through participants’ experiences and concerns within not only these and other teachers’ test-related meanings, effects and interrelationships in the specific school site of testing, but also within the broader contexts: relevant research and literature, and educational and political debates surrounding standardised literacy testing and the value placed on ‘high-stakes’ testing (Carabine; Fairclough, 2001; Scheurich & McKenzie, 2005; Taylor). Further, this chapter adopts the terminology and understandings employed by Gee (1999, 2005, 2011) regarding ‘little d’ and ‘big D’ discourses, although all discourses mentioned in this chapter are the latter type but are not capitalised. This chapter also adopts the terminology and understandings employed by Freebody and Wyatt-Smith (2004) and Wyatt-Smith (2008) regarding the ‘system’ and ‘site’ (or ‘local’) of education debates.

6.1 Defining the Discourses

The previous chapter presented the details and results of the theoretical coding stage, through which the themes were re-examined and discourses identified. The discourses identified in Chapter Five were made evident through the sections addressing the analysis results; as such, the discourses within this

work are thus grounded in the data and reflect the complexities evident in the responses to the research questions. The sections in Chapter Five identified a number of ways in which these tests are valued and by which their validity is considered, and a number of clear patterns of test-related teacher-student action and interaction. These are: literacy diagnosis; school-wide programme planning; time constraints; political (mis)use of testing; test detail; rules, instructions and expectations; negotiation and flexibility; student concern; testing influences; test positivity; and test negativity.

These various identified areas, illustrated and explained in the previous chapter, present a picture of the complexities and compromises faced by those involved in the school site and classroom administration of standardised literacy testing. They show that the participants located themselves within the testing phenomenon in variously accepting and critical ways, identifying strengths and weaknesses in testing as they experienced it, and developing strategies of coping with the testing situation to enable the most accurate representations of their students' abilities in order to make best use of the resultant data.

When examined together for what they say about participants' experiences regarding testing processes and practices, the results of the coding analysis and research questions highlight that teachers' administration of standardised literacy testing happens at the intersection of competing demands and desires, namely, system demands for testing uniformity and site realities of teaching reflexivity. Discursive tensions, resultant of this (dis)connectivity in standardised literacy testing, were evident in participants' explorations of their

testing experiences and in actions and interactions observed in test sessions, exemplified through the two previous chapters.

The two dominant discourses, those of (system demands for) uniformity and (school site realities of) reflexivity, are identifiable through the results of the coding analysis presented and detailed in chapters four and five. These dominant discourses highlight the key patterns of struggle, acceptance and response regarding testing as noted through the data. It is these dominant discourses that are the chief focus of this chapter, with the discourse analysis findings contextualising the dominant discourses within the broader corpus of literature and research that address the topics.

6.2 Dominant Discourses and Discursive Tensions

6.2.1 Discourse of Uniformity.

The Discourse of Uniformity is the overarching patterns of implicit and explicit devices and techniques by which the system-level (Freebody & Wyatt-Smith, 2004) communicated and demanded that standardised literacy testing was expected to be administered and test actors behave. Although evident across all five themes, this discourse is particularly identifiable through the key patterns that constitute themes 1, 2 and 3.

The Discourse of Uniformity is termed ‘system’ as demands and expectations of uniformity in testing originated through school hierarchies, within, outside and above the (school) site level, from demands that stemmed from state and Federal governments and Departments of Education, and the principal’s translation of extra-school pressures and system demands into the site level hierarchy (Mills, 2008). System demands for uniformity in testing

are legitimised and strengthened by media and public acceptance of government noting of the importance of uniformity, standardisation and essentially accountability at the (school) site-level of such testing (Bamford, 2010; Davies & Bansel, 2007; DeBray-Pelot & McGuinn, 2009; Hamilton, 2003; Hursh, 2007; Ognyanova, 2010; Soler, 2002).

The Discourse of Uniformity fulfils a system and a site need, as governmental, departmental and school hierarchies depend on uniformity within testing in order to make use of test data, and discursive compliance and continuity through school hierarchy is necessary to ensure this uniformity (Freeman, 2009). The use of test data for literacy diagnosis, validated against teachers' assessments, for individual student and school-wide literacy programme planning, relies upon accurate and valid test data gained through comparable, uniform testing conditions; assessments must be valid along different criteria and purposes in order to be relevant for use at the school compared with the system, administrative level (Freebody & Wyatt-Smith, 2004; Whitehead, 2008; Wyatt-Smith & Campbell, 2002).

System demands for uniformity within standardised literacy testing are evident through demands and expectations upon test actors at the school site at the planning, organisation, administration and follow up stages of testing. Test actors were required to engage in uniform, identical tasks (most particularly the tests themselves), during which they were also held to a uniform standard of behaviour and expectations. As such, test actors are variously positioned as recipients of already-made decisions at all stages of standardised literacy

testing, and react variously (Craig, 2010; Freeman, 2009; Gawlik, 2009; Kerin & Comber, 2008). Such decisions include:

- Who would be tested: all students of a particular year level;
- When they would be tested: a particular week, not a time selected by teachers;
- Predetermined content of the tests: spelling, writing/composition, and reading/comprehension; the list of spelling words, topic of composition, and so on;
- The materials and support available for students during test sessions: permitted pen, pencil, eraser, but only minimal teacher or teacher aide assistance, and no peer interaction;
- The type and mode of data returned: the presentation of test results, relevance for syllabus or curricula, time of year of results' return; and
- The role of teachers during testing: administrators and facilitators, recipients of results not necessarily helpful for teaching.

All students were given the same tests, expected to comply with one set of instructions, rules and demands, and expected to engage with the test tasks under a particular uniform set of conditions. Compliance with testing and acceptance of various benefits from such uniformity indicates an acknowledgement of the validity and necessity of system-level demands for uniformity (Freeman, 2009; Gawlik, 2009; Willet & Gardiner, 2009). Nonetheless, limitations in transposing system uniformity into the school site without consideration of student test experiences, ability levels and influences upon test sessions was a common concern among participants. A system-level

expectation or demand for strict uniformity over testing was presented as potentially ignoring or devaluing different student needs, resultant of distinct or diverse behavioural or learning levels (Merrett, 2006).

System-level demands for uniformity appeared in opposition to and irrespective of teachers' professional experience regarding testing, classroom administration, student management, needs and learning levels (Hill, 2005; Lingard, 2010). This is so because teachers are provided with rules and instructions with which they are expected to comply, and which do not allow teachers the autonomy to make situated decisions about best ways to administer test tasks and sessions, sufficiently encourage students into completing tasks to a high standard, explain to students what to do, how and when.

6.2.2 Discourse of Reflexivity.

The Discourse of Reflexivity is the patterns of implicit and explicit devices and techniques by which test actors explored and demonstrated their autonomy within the school site. This autonomy entailed a rejection of system-initiated demands for uniformity of teaching and testing approach and, instead, a situation-specific reflexivity that considers the various site-level needs and demands at site level. This discourse is particularly identifiable through the key patterns that constitute themes 1, 3, 4 and 5, although as with the Discourse of Uniformity this dominant discourse is evident across all five themes.

The context of this study employs the term 'reflexivity' as denoting teachers' (and others') capacity to make and execute situation-responsive

decisions. Teacher reflexivity draws on professional autonomy at making site-level decisions, utilising prior experience and with the awareness that they work within a dynamic and fluid classroom environment and relationships (White-Smith & White, 2009; Wilson, Croxson, & Atkinson, 2006). A system expectation of site-level uniformity inconsiderate of participants' perceived and experienced school site realities was a core factor of participants' criticisms as to the value and validity of standardised literacy testing, as uniform rules, tasks, conditions and comparisons served to exacerbate and compound testing pressures for school site-level test actors (Freebody & Wyatt-Smith, 2004).

Site reflexivity positions test actors as able and willing to respond to situations by making use of their experiences and knowledge through standardised literacy testing processes and practices. Educators argue the benefits of utilising their professional judgement in their own and their students' capacities, and how to best build on them (Beutel, 2010). The Discourse of Reflexivity is evident through teachers' decisions to act against set rules, to administer and utilise tests and test results in ways that serve their professional purposes and reflect their professional judgements, and to articulate their resistance of demands for uniformity at all stages of testing and their reasons for resisting demands for uniformity.

Careful consideration of the classroom management and test administration approaches, in order to employ that which would have best effect on task implementation, is clearly important for teachers (Garrett, 2008; Hoy & Weinstein, 2006). Use of humour (Ullucci, 2009) as well as behaviour

negotiation and, on occasion, classroom relocation (Kulinna, 2007/2008) as preferred classroom management approaches indicate teachers' employment of various strategies in response to student misbehaviour, more positive than punitive, in order to maintain working classroom relationships while also altering student behaviour (Van Petegem, Creemers, Rossel, & Aelterman, 2005). Positive teacher-student relationships are positively related to students' educational outcomes (Beutel, 2010; Den Brok, Levy, Brekelmans, & Wubbels, 2005; Parker, 2006; Thwaite & Rivalland, 2009). Flexible classroom teaching is additionally important with low literacy students, given the possibility of undermined student confidence caused by peer bullying or specific student needs (Ballis, 1999/2000; Emmer & Gerwels, 2006; Pianta, 2006; Richel, 2005).

Within teachers' reflexive and considered classroom approach is an acknowledgement that negotiation between teachers and students is a necessary element of the learning environment (Park, 2008), that all classroom environments differ in tone, interaction and behavioural norms (Walker & Adelman, 1976) and that even in constructive classrooms, teachers and students can hold different understandings of these norms and expectations (Boomer, 1998; Eisenhart & Borko, 1993; Furlong, 1976; Reeve, 2006).

6.2.3 Discursive tensions.

The dominant discourses of system uniformity and site reflexivity played out in opposition to each other. Uniformity demands play out in the school site of testing with its demands for standardisation across test-related processes and

practices, against which teachers' reflexivity was positioned. The school site test actors, with their attempts to engage with standardised literacy testing on their own terms, often challenged such uniformity, although nonetheless acknowledged beneficial outcomes of testing uniformity such as validity and comparability of test results.

The distinct demands and needs at the system and the (school) site levels (Freebody & Wyatt-Smith, 2004; Williamson & Myhill, 2008; Wyatt-Smith, 2008) of standardised literacy testing create space for tension for school test actors. At the system level, the authority held by government and by those holding senior positions in the education hierarchy beyond the school site meant that participants were forced to submit to externally-set rules, such as what Olsen and Sexton (2009) described as "applying strong pressure for school personnel to conform" (p. 9). At the (school) site level, participants admitted to selectively breaking rules they considered inappropriate for, and inconsiderate or unappreciative of, their school site realities. Participants suggested that those outside of the school site had less understanding and appreciation of the school site than themselves, but nonetheless held the autonomy to demand uniformity (Grant, 2009; Webb, 2005).

In this way, although in most situations holding power and autonomy in the classroom (Delamont, 1976), teachers held limited authority and power through the various testing processes, practices, sites and levels, as an extension of the educational power hierarchies within which they operated and the blame given to teachers for their role in apparently failing literacy standards (Blanchard, 2003; Florian & Rouse, 2001; Lingard, 2010; Snyder,

2008, 2009). TT's comment that the tests' authors could or should not be involved in class teaching indicated his opinion about the disconnect between test expectations and demands, and realities of the school site. In spite of perceived issues within testing and accountability measures, though, participants did accept and appropriate some practices (Gawlik, 2009; Warne, 2006; Webb, 2005).

The inherent underlying tensions between what is here termed system uniformity and site reflexivity are at the core of arguments regarding the concept of and responses to 'difference'. An identification of comparative academic and behavioural differences across schools explains teachers' prioritising of professional expertise and reflexivity over uniform treatments and considerations of students (Akiba, LeTendre, & Scribner, 2007; Alvermann & Reinking, 2007; Lee, 2010; Lee & Wong, 2004; OECD, 2010).

At a time when classrooms are filled with diverse compilations of students, and when the futures that these diverse students are headed toward are more uncertain than ever before, a dominant and powerful set of representations ill suited to our current context continues to thrive within these accountability explanations. (Woods, 2007, p. 2)

This concept of difference is an issue at the system level rather than only the site level of education, as it is often perceived as a problem to be solved, considered in policy and practice and taken note of in and outside the classroom as a potential strength as well as a factor in some unequal educational outcomes (Beane & Apple, 2007; Brooks, Maxcy, & Nguyen, 2010). Participants' professional struggles in reflexively and considerably

responding to student differences within system pushes for testing uniformity demonstrate an awareness of the effects of such differences upon educational outcomes (Comber & Kamler, 2004; Comber & Nixon, 2009), and indicate a conflicted valuing of the uses of test data in highlighting how teachers (and others) can best identify and address students' literacy problems.

The classroom effects of system demands for uniformity within testing show the conflict between the two dominant discourses, with varying acceptance and rejection of calls for uniformity from both the system and site levels. Negotiation and reflexivity in the classroom and non-testing or nonuniform situations has been argued to improve student understanding and overcome or minimise student discouragement (Beutel, 2010; Collidge, 2001). Within the test situation, however, demands for uniformity in administration approach, rules and tasks (as stemming from broader trends in education) can leave teachers – who otherwise enjoy a position of classroom authority and comparative autonomy – in the position of balancing their professional considerations against set instructions and guidelines (Lingard, 2009), and such resultant pressures can have consequences for teachers' work lives and relationships with students (Churchill, Williamson, & Grady, 1997; Valli & Buese, 2007; Williamson & Myhill, 2008).

Uncertainty and inconsistency regarding the most appropriate approach to classroom test administration and behaviour management can be interpreted or understood as an acknowledgement of broader student problems. A perceived student culture and extraneous issues and problems can influence classroom order and student behaviour, a reminder of behavioural challenges

with particular student groups and the importance of minimising student disruptions to maintain classroom order. A situation-reflexive, considered approach to classroom control and lesson administration can facilitate required test activities, in cases of issues related to school culture, the nature of the testing tasks and the students' testing (in)experience, but can also encourage teachers to act in such a way during test sessions that (negatively) affects test results (Abrams, Pedula, & Madaus, 2003; Braun, Zhang, & Vezzu, 2010). Total rejection of uniformity of standardised literacy testing would undermine the benefits of such educational initiatives as standards-based literacy testing (Wyatt-Smith & Klenowski, 2010).

The use of test results as a justification for judgements and treatments of particular schools can correlate with the perception that enforced uniformity, through apparently inappropriate and unrealistic test rules, processes, practices and follow-up, facilitate teacher deprofessionalisation (Ball, 2009; Harris, 2005; Hodgetts, 2010). In this regard, system demands for testing uniformity are indicated as working towards two particularly rejected aims. The first of these is its perceived encouragement of teacher deprofessionalisation; the second is the reliance upon testing uniformity for decisions to be made that enable what these participants variously indicated as potentially damaging and deceptive comparisons of students, schools and teachers (Bernal, 2005; Demaine, 2005; Hursh, 2007; Kohn, 2000a, 2000b; Lingard, 2010; Merrett, 2006; Sacks, 2000). Although accepting the value and usefulness of school-wide test data, different degrees of ambivalence about even value-added results as used to compare performance highlights the lack of consensus about how

test data ought best be used (Gorard, 2006; Jensen, 2010b; Kelly & Downey, 2010; Kelly & Monczunski, 2007).

In spite of these identified issues apparently resultant of the Discourse of Uniformity, discursive tensions characterise participants' positioning in rejection of testing uniformity. Participants' acceptance of particular elements of test uniformity that reflect 'best practice' reflected arguments in favour of testing processes, in acknowledging the need for and use of gauging student literacy levels, such as to check against internal results, highlight disparities between student performance on normal assessment tasks and on standardised literacy tests, and to assist school-wide planning; the experience that testing gave students in performing under examination conditions, which they otherwise did not have; and the 'value' served by uniformity in testing, which provided more 'valid' results that could be used with more certainty of their credibility, for those so inclined (Afflerbach, 2009a, 2009b; Cohen, 2010; Stringfield & Yakimowski-Srebnick, 2005). The discursive tensions between system demands for uniformity and site realities of reflexivity then arguably represent the broader debates surrounding standardised literacy testing, which themselves are contradictory, conflicting and lacking in consensus.

6.3 Effects of the Discourses

As a result of system demands for uniformity, standardised literacy test data can be presented as uniformly collected, generated and quantified, and therefore can be considered valid for use as reliable and comparable statistical measures of student literacy levels and the identification of high- and low-

performing schools and states/territories (Rich, 2000). System demands for uniformity within and through the school site of standardised literacy testing serve a purpose outside of education system and site levels, for the wider social context. Guarantees of uniformity allow the public, governments, test actors and interest groups to trust not only what happens during testing, but also that what happens during sessions results in more valid and comparable data. Testing uniformity further allows the public to trust what teachers do, to believe that their children are being treated in the same way as their peers across the school, state/territory and country. This Discourse of Uniformity can serve the purpose of relieving parental anxieties about the quality and equality of their children's education; allowing the public to more confidently compare student, school, and state/territory test results; and making teachers and schools more publicly accountable.

Uniformity in testing validates the value and utility of quantitative educational measures, in turn serving to increase the amount and supposed reliability of that data upon which decisions about educational comparison and competition can be made. In this context, educational comparison and competition is shown through test data being used as a basis on which literacy results are compared across schools, in order to compete for students, government funding, and public estimation (Crow & Weindling, 2010; Gordon & Whitty, 1997; Hill, 2005; Wilson, Croxson & Atkinson, 2006). Relatively minor differences in test results are at times over-emphasised, further justifying the call for improvements (Levin, 2004; Stringfield & Yakimowski-Sreblick, 2005; Torgerson, 2006).

Teacher uncertainty about making sense of test data combined with pressure for success have been identified as leading to arguably unethical teacher behaviours of subverting test rules, such as coaching for better results and exclusion of students predicted not to improve school ratings (Abrams, Pedulia, & Madaus, 2003; Booher-Jennings, 2005; Patty, 2010; Wilson, Croxson, & Atkinson, 2006). Confusion as to how to interpret test data could be addressed through specialised professional development (Cumming, Wyatt-Smith, Elkins, & Neville, 2006; Wyatt-Smith, 2008), and uniformity in testing, if it assists in the generation of valid information on student literacy standards, could contribute to the detail teachers utilise in reflexively addressing student literacy needs, and thus serve a valuable purpose (Kerin & Comber, 2008).

System demands for, and ensurance of, uniformity in standardised literacy testing provide clear data for school comparisons which highlight weak links in the system needing to be improved or held accountable (Lingard, 2010). However, it is the test-related media and political pressures that can be associated with this consequence of test uniformity and use, and that enable teachers to be judged according to their students' performance, which does not necessarily improve as a result of such high stakes testing (W. B. Black, 2004; Doecke, 2007; Hodgetts, 2010; Mills, 2008).

Even for teachers keen to improve students' literacy levels, system pressures upon the school site to always improve student outcomes and teach in certain ways, expressed in the form of policy and dictates, can make teachers feel constrained and non-autonomous in what they do (Achinstein, Ogawa, & Speiglmann, 2004; Beck, 2008; Clarke, 2001; Coldron & Smith, 1999; Parr &

Bellis, 2006). The perception of a public and system unawareness of the lack of resources, time and funding, with which participants presented they are regularly required to work, is not only frustrating for teachers but also affects teachers' perspectives and what teachers can do (Christensen, 2010; Churchill, Williamson, & Grady, 1997; Goodwyn, 2000/2001, 2004; Williamson & Myhill, 2008). This in turn has been found to lead teachers to be increasingly controlling with students over factors they could still control (Pelletier & Sharp, 2009). The popular welcoming of competition of and across schools, reliant upon popular assumptions and considerations of school accountability and competition, is representative of further educational and political issues for teachers (Goodwyn & Findlay, 2002).

The participants presented the main reason for their desire for reflexivity within the testing processes as the disregard given to their professional expertise by the demanded uniform approach to administering and utilising the tests. Dominant educational discourses identify “the ‘good’ teacher [as] not only a ‘competent craftsman’ ... but also one who is submissive to authoritative discourses” (Honan, 2010, p. 190), against which “even quite small, seemingly inconsequential acts of resistance or non-compliance can appear to be heroic and radical acts of agency” (p. 190).

Site reflexivity might enable test sessions to be administered in an inclusive way, with more students involved in tasks and contributing to test outcomes and generated data. In this way, the broader order of standardised literacy testing and network of educational and political practices stemming from or dependant upon standardised literacy testing do benefit from site

reflexivity in testing, even only from its consequence of allowing more data to be generated.

Teachers' reflexivity in response to demands for uniformity and standardisation highlights perceptions of teacher disempowerment and deprofessionalisation as a direct result of system demands for uniformity within the test situation. Demands for uniformity stemming from the broader educational system and seeking to shape behaviours are presented in test rules and procedures establishing that teachers (as test administrators) are unable to take control of the testing procedures – act reflexively or autonomously – without being considered to have circumvented or broken set rules and invalidated test data.

6.4 Absences and Silences within the Discourses

Absences and silences within a dominant discourse are important to question because such voices are hidden for a reason. As argued by Taras (2007):

ignoring the challenge and refusing to take up the gauntlet is perhaps the best and most efficient means of silencing the opposition. If a voice is not heard, no matter how potentially loud or vociferous, then the voice does not exist. (p. 67)

The Discourse of Uniformity effectively silences alternative discourses from gaining legitimacy: test actors' concerns about the Discourse of Uniformity, their reactions and alternatives to the Discourse of Uniformity (Reflexivity), and their reasons for adopting alternatives to the Discourse of Uniformity. This is evident through the Discourse of Uniformity's strategy of overriding teachers' situation-reflexive decisions or responses, through dictating rules and

valuing identically-obtained data, and through broader system silencing of research that challenges popular assumptions about the value of such approaches, including high stakes literacy testing (Freebody, 2007). Further silenced in the Discourse of Uniformity is appreciation of changes in teachers' work situations and expectations, which not only increase external and policy-related pressures on teachers but also mean teachers are required to respond to varied classroom environments (Churchill, Williamson, & Grady, 1997; Williamson & Myhill, 2008).

Alternative voices to the Discourse of Uniformity, site and classroom nonuniformity and autonomy are silenced and devalued through increasingly uniform and standardised educational regimes and practices; what is valued through the Discourse of Uniformity is chiefly a comparable uniformity of conditions, input and outcomes, which can be seen as beneficial too (Phelps, 2005). The silencing of teachers' voices and experiences within the testing phenomenon could be addressed or redressed by increased system-level respect for teachers' expertise in how to test or gauge students' literacy levels, just as a valuing of teachers' assessments alongside summative, standardised tests could provide a more realistic representation of literacy levels (Brown, McCallum, Taggart, & Gipps, 1997; Zyngier, 2009).

System pushes for testing uniformity are legitimised by those within a position of power and influence within education, with a strong 'voice' (government, Department of Education, media). In the process of legitimising the dominant system voice, those demanding strict testing uniformity perpetuate the public perception of the alternative voices as being opposed to

such measures because of teachers' fear of being held accountable (Phelps, 2003) and through the silencing of teachers' professional expertise (Beck, 2008; Prashad, 2006; Warmington & Murphy, 2004). Alternatives to system uniformity can also be dismissed with reference to their 'ideological' standpoint: the labelling of the alternative as ideologically-motivated, which diminishes the value or validity of their argument.

Absent and silent within site pushes for reflexivity is a genuine acknowledgement of the value of uniformity. This silence is highlighted and problematised through the inconsistent acceptance and rejection of the Discourse of Reflexivity, which indicates that neither of the two dominant discourses was wholly accepted at this school site. Further absent within and across both discourses, too, is a clear voice of the benefits and consequences of such measures as high stakes testing upon students (Busher, 2006; Cizek, 2009a, 2009b).

Concerns that value of testing is undermined for those at the school level through issues of accessibility of data are not isolated to this research (Gawlik, 2009), and the silencing of such calls suggests a system privileging of their access to data or design to implement processes rather than empower teachers to utilise such information directly and quickly.

6.5 Key Findings

The discourse analysis of the collected data (Carabine, 2001; Fairclough, 2001; Taylor, 2001a, 2001b), utilising the initial results of the constructivist grounded theory coding analysis (Charmaz, 2000, 2001, 2005, 2006; Charmaz &

Mitchell, 2002) and the research questions, has contextualised the research in the broader network of related practices, power/knowledge structures, social implications (effects), representation and contestation. In so doing, from the discourse analysis can be drawn five particularly salient findings which connect the examined phenomenon and case study to the extant literature on this topic. These findings are variously related to each other through their effects upon those involved (in the research, and in the testing phenomenon in a broader sense) and through the ways in which they are enacted and to which test actors respond.

Finding 1: Participants demonstrated varying opinions as to the value and validity of standardised literacy testing and test results, in general and regarding certain students in particular, which correlated with varying approaches to testing.

The participants expressed an acceptance of the needs, especially the political and monitoring needs, for such testing, but identified specific ways in which the tests and testing were inappropriate for their own and their students' needs. Participants' aired frustrations about comparisons made through and of test data, about a public connection drawn between student test results and teaching quality, and the ways testing took place, contrasted with their (particularly T2's and P's) acceptance of the need and value of standardised literacy testing. This finding counters claims made about educators by Phelps (2003), who argued that educators who dislike or misuse standardised literacy testing do so out of a fear of being held accountable for their students' and their own poor performance.

The participants did not claim that standardised literacy testing is unavoidably flawed and invalid, but that it is flawed and invalid in some ways that could and should be addressed. This finding both challenges and complements those of Johnson, Johnson, Farenga and Ness (2008), whose research found overly negative teacher perceptions of standardised literacy testing. Particularly, this research suggests that in spite of the negative aspects of standardised literacy testing, these participants still identified benefits and uses, value and validity, of testing.

Concerns about the language and content of the tests and what is considered to constitute student intelligence appear to confirm arguments made by Gordon (1999), Hursh (2008), and Snow, Porche, Tabors and Harris (2007). Those authors raised the issue of popular beliefs regarding the importance of questioning students' social and cultural capital as reflected in the tests and by the test authors.

Participant emphasis of issues pertaining to language, literacy and intelligence within standardised literacy testing contributes to research undertaken by Matoush and Fu (2008) and the OECD (2000) regarding language users' employment of different forms of language. This finding supports arguments for more sympathetic consideration of realities of student language use and how such usage should inform the development of standardised literacy testing for greater validity.

These participants were critical of what they presented as unhelpful delays in the return of test data to them. The tests observed in this study were administered in the final week of July, and results were returned to the

principal towards the middle of November of the same academic year - a gap of over three months. The resultant lateness in the academic year of the release of the test results and data to schools meant a period of no more than six weeks remained in the school year for teachers to adjust their syllabi or teaching schedule. The participant argument that literacy data should be made available considerably sooner to school and teaching staff reflects similar arguments made by McCaffrey and Hamilton (2007) and Ravitch (2003) in their studies into the school use of such test data.

The data clearly show that participants considered standardised literacy testing to be not always appropriate for all students and not reflective of all student literacy levels. The idea that these tests should not be considered in isolation or overvalued, due to the incomplete data provided by the testing and due to the influences upon students that affect their results, is corroborated through research by Gipps (1999). That research suggested that the reality of standardised literacy testing and reporting of test data means that statistical outcomes and results neither explain student results nor differentiate between students whose low results are due to lack of academic ability, educational opportunity, or teacher ability.

Student test results in recorded quantitative and statistical form cannot also inherently provide the kind of evidence required for qualitative interpretation of student scores; participants expressed their concern about this issue. These participants' concerns expressed in discussions of students' socioeconomic status and articulated perceptions about a poor work ethic feature of student culture are similarly explored within the research and

academic literature. Correlations have consistently been found between students' educational performance with the affluence of school area, parents' educational qualifications, employment levels and socioeconomic status (Anderson, Medrich, & Fowler, 2007; Gardner, 2007; Hirsch, 2007; Levin, 2007). These data and the findings are contextualised by a consciousness of public calls for equitable school funding and support. The findings signal aspects of the lived school experience that could contribute to alarming disparities between some student outcomes, especially disparities linked to socioeconomic status and those associated with state/territory comparisons in standardised literacy testing results, as recently maintained and upheld in NAPLAN reports (ACARA, 2010b, 2010c; Ladwig, 2010; McGaw, 2007; Nguyen, 2010; OECD, 2009, 2010; Wyatt-Smith, 2008).

Frustrations and criticisms based on participants' perceptions that standardised literacy testing was inappropriate for particular students, that test data are possibly invalid, misused and over-relied upon, indicate or point to a possibility that these participants did not consider equal opportunities to learn existed for students and that testing and judging students should not proceed as if those conditions did exist. This finding of the research too corroborates other produced work (Gee, 2003; Johnson & Johnson, 2006; Kozol, 1992, 2005; Newman, 1995; Nieto, 2005; Popham, 2001, 2007; Stanley, 1995).

As such, concerns, particularly verbalised by TT and P, about the effects of school catchment and student culture upon student literacy outcomes are echoed in the broader literature. Examples from the data include references to students' average socioeconomic status, comparisons between their school

site and other schools, the expressed belief that the tests are often too difficult for their student cohort, and discussions of students' experiences with police and social workers. Participants' concern about the effect of their students' socioeconomic status and other factors upon students' test results supports, and is supported by, the significant body of research claiming school and non-school factors are more influential upon the test performance of minority and low-socioeconomic status students than they are upon other students (Anderson, Medrich, & Fowler, 2007; French, 2008; Gardner, 2007; Gipps, 1999; Gordon, 1999; Hirsch, 2007; Johnson & Johnson, 2006; Messick, 1999; Newman, 1995; Popham, 2001, 2007; Snow, Porche, Tabors, & Harris, 2007; Stanley, 1995; Tankersley, 2007; Tharp, 2008). Reflecting Gee's (2003) arguments about the importance of adequately recognising student opportunity to learn, participants expressed concerns about students' personal and educational context that shaped their interpretations of these students' opportunities to learn. Student contextual learning opportunities thus emerged from this research as a significant issue for testing.

Participants considered standardised literacy test results and data as providing incentives for, or necessitating, syllabus and curriculum change. Participants, although not T1, evidently believed standardised literacy testing could assist in the structuring and administration of learning topics, which in turn could affect student learning and learning outcomes. This finding also supports and is supported by the work produced by FairTest (2008), McCaffrey and Hamilton (2007) and Messick (1999) who also presented such claims.

In relation to participants' opinions of students, issues affecting and possibly undermining potential benefits of the tests were identified in the research. The tests were often presented as possibly too challenging, and subsequently unappealing, for this group of students. TT in particular presented the tests as so excessively difficult for students being tested that the tests' content served to actually skew student test results and data. As a result, the tests' educational value and potential for use were undermined, according to TT. This aspect of participant concern is addressed by Johnson and Johnson (2006) and Johnson, Johnson, Farenga and Ness (2008), who argued that such tests would be considered more appropriate by teachers if they allowed for consideration of students' different levels of literacy development, rather than allowing these to act as a barrier for students' educational success. The literature similarly reflects participant frustrations at what they considered to be inappropriate use made of tests, such as in some school comparisons. These data appear to confirm Van Blerkom's (2009) claim that teacher frustration over (in)appropriateness of school comparisons and test use can undermine for them the value of standardised literacy testing.

Finding 2: Participants perceived an increasing trend towards teacher deprofessionalisation, as a direct result of such measures as high stakes teaching, political use of test data and school comparisons.

Irritation was expressed of the particular use made of standardised literacy test results in judging and potentially discriminating against particular schools. A related frustration pertained to a perceived increasing use of media sensationalism based on the use of literacy standards data. Ire at punitive

measures directed at schools, teachers and students, and dissatisfaction that the learning conditions and contexts of their students are ignored in comparisons of school or district literacy results, confirms and supports a considerable body of educational research that reaches similar findings (Ambach, 1999; French, 2008; Hursh, 2008; Johnson & Johnson, 2006; Johnson, Johnson, Farenga, & Ness, 2008; Messick, 1999; Phillips, 2004; Popham, 2001, 2004; Rohlen, 2002; Tankersley, 2007).

Calls for 'value-added' test data analysis, such as by the Grattan Institute (Jensen, 2010a, 2010b), challenge and attempt to redress this issue of neglected consideration of students' contexts in standardised literacy testing and test comparisons. These data further show the issue as a major factor for these participants in their perception of the value and validity of standardised literacy testing and their consequent approach to test administration and follow-up.

The data connected perceived misuse of testing and test data with accountability issues. Participant concerns about teacher deprofessionalisation highlighted one potential outcome for them of accountability demands, which TT and T1 considered a likely development in the near future of literacy testing. The data show that the participants correlated high-stakes testing with accountability measures, school league tables and 'merit-based' teacher pay, all of which the participants identified as issues of concern, consistent with current research knowledge (Hanushek, 2001; Hursh, 2008; Johnson & Johnson, 2006; Johnson, Johnson, Farenga, & Ness, 2008; Kennedy, 2008; Rothman, 2001; Tankersley, 2007; Tyack & Cuban, 1995).

Baines (2007) and Baker (2007) demonstrated the international relevance of these concerns: many education systems require schools to submit to identical, ‘one-size-fits-all’ tests that reveal differences in students’ opportunities to learn. Their research also showed the extent to which teachers in numerous countries are subjected to threats of sanctions and promises of potential for rewards (EdSource, 2002). This research project’s findings show that these participants were deeply concerned about possible moves towards what they perceived as constricting educational standardisation.

Finding 3: Participants perceived a lack of professional respect given to teachers’ expertise in test construction, expectations and use.

The data highlight participants’ numerous, varied and specific concerns as to the administrative and political uses of standardised literacy tests, which in turn suggested their frustration that potentially problematic measures are privileged and valued at the system level instead of teachers’ own site-valid forms of assessment (Freebody & Wyatt-Smith, 2004). Participants identified that issues within standardised literacy testing rendered a reliance on the tests’ data misplaced, just as their identification of time frame issues indicated their similar opinion at a reliance on timely usage of results.

This finding challenges claims about testing such as those made by Van Blerkom (2009) who claimed that standardised literacy tests are not just “technologically and methodologically very sophisticated” but also “obtain much higher reliability and validity” than similar tests created by teachers (p. 189). This finding supports the arguments posited by Popham (2001) that

beliefs of test validity and objectivity are misconceived and that standardised literacy testing does not fully allow for an appreciation of teachers' experience and knowledge.

Participants, particularly TT, were desirous of increased teacher involvement in test design and decision-making. This finding has relevance to consideration of teacher involvement in the setting of the literacy standards, which is promoted in much of the literature and research regarding standardised literacy testing (Johnson, Johnson, Farenga, & Ness, 2008; Nelson, McGhee, Meno, & Slater, 2007; Rhodes, 2007; Shanker, 2007).

According to these data, understanding of the educational merit of high stakes, standardised literacy testing are contingent on a range of factors such as classroom actors' interactions and relations, environmental influences, and perceptions and prejudices regarding the tests, of which teachers are aware and are able to consider in their own assessments. The data and broader literature suggest that increased teacher involvement in test design and writing could help to improve tests. Research produced by Tucker and Coddling (2001) supported delegating important professional responsibility (back) to teachers, for example by increased involvement in test design and decisions regarding test administration and use of data. Tucker and Coddling argued that teachers need to be supported, supportive and involved if they are to administer these tests, rather than be expected to follow rules set by others and considered responsible for the outcomes.

The findings of this and related Australian education research mutually reinforce calls for the practical acknowledgement of teachers' experience,

professionalism, high levels of training and expertise (Everson, 1999; Hawkins, 2008; Hursh, 2008; Johnson & Johnson, 2006; Luijten, 1999; Phillips, 2004; Popham, 2001; Robertson, 2010; Stiggins & Duke, 2008). At the international level, DeCoker's (2002a, 2002b), Fuhrman's (2002) and Russell's (2002) international comparisons of education systems and processes similarly validate these findings advocating direct teacher involvement in all stages and aspects of education.

Participants' desire to be involved in various aspects of education policy and practice, and their apparent willingness to work towards educational improvements through the utilisation of their skills, expertise and experience, supports arguments that teachers should be much more involved in educational research. This suggestion accords with conclusions reached by a number of education researchers (Britzman, 1995; Felson, 1982; Giroux, 1995; Rhodes, 2007; Tharp, 2008; Tucker & Coddington, 2001). The finding that participants valued that which was relevant to their work as teachers and to their school context also supports arguments expressed by Brewer and Goldhaber (2008) and Fusarelli (2008), that education research should be directly relevant to schools in order for it to be utilisable by principals and teachers. Participant anger at apparently misinformed and unhelpful judgements of their and their students' performance supports research by Bracey (2007) and Rotherham (2008) who urged that teachers, educators and students must reclaim the power and potential of the classroom if they wish to avoid perceived media-dominance over what is disseminated about education and schools.

Finding 4: Participants differently interpreted test administration rules and guidelines, according to perceived student needs, predicted behavioural issues, and their intended use of test data.

A clear finding from the data is a participant inconsistency with test rules when planning, administering and following up on standardised literacy testing, particularly through selectively ignoring test rules during test sessions. The data indicate participant awareness that not all students are able to achieve a high or even average result on the standardised literacy tests, due both to the nature of the tests as diagnostic benchmark tests and to the perceived student abilities. This awareness appeared to encourage participant encouragement of their students, which the participants presented as important, to facilitate relatively disruption-free test administration and classroom management. The data suggest that such encouragement of students through the standardised literacy test situation was additionally important because of the individual, high-pressure testing environment; the teacher and teacher aide could still choose to interact with students, albeit with restrictions of the kind of support or interaction provided.

The data suggest that participants employed this rule inconsistency to help them maintain classroom order and demonstrate empathy for struggling students. Abrams, Pedula and Madaus (2003), Henningfield (2008) and Johnson and Johnson (2006) highlighted claims that teachers resort to 'cheating' when administering these tests, and reminded that these claims are variously expressed by educators, researchers, politicians, media and interest groups. Although admitting to subverting some test rules and justifying this

behaviour in different ways, these participants stopped short of acknowledging that they engaged in any test 'cheating'.

The data demonstrate that teacher approaches to student interactions can greatly affect students' experiences, engagement or disengagement with testing. These findings are similar to those of Snow, Porche, Tabors and Harris (2007), who expanded this argument to posit that teacher-student interactions can shape student understandings and interpretations of school, peer and self. Participant opinions of the importance of fostering encouraging, positive relationships with students, observed during standardised literacy testing sessions and discussed in interviews, seemed part of an attempt to encourage positive student experiences with schooling.

Encouragement during test administration could send the covert message to the students about their abilities and the different expectations teachers had of them, that some students were unable to complete even the simplest of the test requirements. Such behaviour could reinforce the students' lower skills by allowing for differentiated treatment, and students receiving the extra assistance might feel worse if they later discover that their results were lower than the teacher had encouraged them to believe. Participants expressed the belief that encouraging and positive interactions with students would allow students the motivation to work harder during tests; an unintended consequence could be that students have more fear of 'failing' the test and disappointing the teacher, having a huge impact upon students striving for teacher approval and recognition (Johnson & Johnson, 2006).

Finding 5: Participants were sceptical of policy writers' (system) understandings and appreciation of the classroom (site) reality of testing.

This final finding reflects the overarching tensions and perceptions evident within the other findings. The identified issues, both in these data and across the canvassed literature, demonstrate that teachers who administer and utilise testing at the classroom and school site are aware of a number of issues that they consider to be of pressing importance. This suggests that those at the system level are less concerned with the practicalities and realities of classroom testing at the site level than with the uses made of testing for system purposes.

The participants articulated a number of concerns and issues regarding standardised literacy tests, ranging from test-specifics, such as the use of numbering for cloze spaces in the spelling tests and time frame issues, to wider issues, such as the broader uses of the test data for government and accountability purposes. In this way, the data and the findings highlight what participants presented as a gulf between standardised literacy testing in theory and in practice, wherein the test reality at the school level is neither considered nor appreciated by test authors and governments when devising the standards, testing documents, guidelines, timeframes and repercussions. Such findings expand the literature on standardised literacy testing, providing further justification for test authors and politicians to consider the classroom effects and broader consequences of standardised literacy testing, as well as the broader social environment outside of the classroom (Felson, 1982; Giroux, 1995; Macedo, 1995; Phillips, 2004).

Chapter Seven: Conclusion

Introduction

The first section of this concluding chapter, 7.1 Research and Limitations, briefly revisits the research aims, research questions, analysis approaches and data sources, and addresses the research limitations. The second section, 7.2 Key Findings, summarises the five findings of the combined coding and discursive analyses. The third section, 7.3 Recommendations, explores contextualised suggestions for future research and education policy and practice. The fourth section, 7.4 Research Contribution, locates the study in its context, clarifying its importance and relevance for researchers, practitioners, policy writers and the public.

7.1 Research and Limitations

7.1.1 Overview of the research.

This research undertook to examine the topic of teachers' administration and experiences of standardised literacy testing through a combined constructivist grounded theory (Charmaz, 2000, 2001, 2005, 2006; Charmaz & Mitchell, 2002) and discourse analysis (Carabine, 2001; Fairclough, 2001; Taylor, 2001a, 2001b) approach. Two research questions were addressed in the research: 'How do teachers perceive and interpret the value and validity of standardised literacy testing?' and 'What patterns of action and interaction characterise standardised literacy testing?' In the context of this research,

‘value’ referred to importance and usefulness, and ‘validity’ to accuracy of test results as student literacy indicators.

A qualitative, case-study approach was employed in examining this topic, enabling the research to focus on one school site in order to highlight the complexities and nuances in the experiences and perspectives of one small group of teachers. The influence upon this research of postmodernist and poststructuralist theory (Carroll, 2002; Lather, 1991; Weedon, 1987; Whipp, 1998) communicates the researcher’s appreciation of multiple perspectives, interpretations and potential meanings. An appreciation of complexity is important because of the nature of the research topic itself, involving identification and exploration of participants’ subjective interpretations.

This research was conducted at one Tasmanian school site, with four participants: three teachers (TT, T1 and T2) and one principal (P). Two class groups were observed as they participated in literacy testing. Participants were interviewed and completed questionnaires; the two observed class groups were also given questionnaires. The observed test sessions and interviews were audio taped and transcribed, and the researcher recorded field notes throughout. Necessary clearance was sought and granted prior to the data collection, from the University of Tasmania and Tasmanian Department of Education. Informed participant consent was sought and granted prior to the researcher’s data collection at the site.

7.1.2 Possible limitations.

7.1.2.1 Methodological limitations.

The involvement of only a small sample of teacher participants at one school site represents a potential methodological limitation in the form of situated knowledge. Generalisability of findings, therefore, needs to be treated with caution. Translating findings and suggestions from this one location or drawing inferences with respect to other contexts or situations, in particular to other school sites, could be problematic.

This is not to minimise or negate the value of the research; an appreciation of methodological limitations does not negate its strengths. This research produced grounded findings, albeit situated in one particular context, and makes no inflated claims to objective, universal results or findings. As Taylor (2001b) argued, however, “*all* knowledge is considered to be situated, contingent and partial ... reality is also inevitably influenced and altered by any processes through which a researcher attempts to investigate and represent it” (p. 319). Rather, the researcher examined multiple data types (interviews; test transcripts; field notes; questionnaires) to more confidently identify patterns, and utilised multiple analysis approaches (one approach to coding, and two to discourse analysis) to more thoroughly scrutinise the data.

7.1.2.2 Instrumentation limitations.

Researcher involvement can affect the research in a variety of ways, such as through the selection of the topic, data generation and collection, interpretation and analysis, and the results, findings and recommendations drawn.

Researcher bias is another element to be considered. It could be argued, for

instance, that the researcher's involvement and investment in education processes and practices could have affected or jeopardised this research in particular ways. All research is shaped by the researcher, however, and none is free from researcher limitations. Therefore, even replication of this study by other researchers could result in findings with different limitations, at the same time contributing to the overall body of knowledge on the topic.

Qualitative research requires the researcher to make meaning of and interpret data, yet meaning and interpretation are not uniform and are very much shaped by personal opinion. It is acknowledged that this further potential limitation could apply through the body of this research. Hatch (2002) referred to the researcher as “making inferences, developing insights, attaching significance, refining understandings, drawing conclusions, and extrapolating lessons” (p. 180). In this research, subjective interpretations were formed with insights and explanations of the “social situation” (Hatch, p. 180) of standardised literacy testing. Here this idea of researcher as both insider and outsider can be considered a potential limitation as well as a source of strength for the produced knowledge.

The researcher in this study, as both a researcher and a teacher of English/Literacy, approached the data and issues arising from contrasting and complementary perspectives. The researcher appreciated the possible difficulties caused by proximity to the data, and employed dual analysis approaches conducted systematically with multiple forms of data in order to gain confidence in the work produced. The case study utilised a combination of a constructivist grounded theory coding analysis (Charmaz, 2000, 2001,

2005, 2006; Charmaz & Mitchell, 2002) and discourse analysis (Carabine, 2001; Fairclough, 2001; Taylor, 2001a, 2001b). This allowed the researcher to feel more confident of achieving the kind of critical, external perspective described by Carabine (2001) as aiding a more thorough examination of discourses, meanings and practices.

7.2 Key Findings

This research produced five key findings, which were presented and contextualised in Chapter Six as:

1. Participants demonstrated varying opinions as to the value and validity of standardised literacy testing and test results, in general and regarding certain students in particular, which correlated with varying approaches to testing
2. Participants perceived an increasing trend towards teacher deprofessionalisation, as a direct result of such measures as high stakes testing, political use of test data and school comparisons
3. Participants perceived a lack of professional respect given to teachers' expertise in test construction, expectations and use
4. Participants differently interpreted test administration rules and guidelines, according to perceived student needs, predicted behavioural issues, and their intended use of test data
5. Participants were sceptical of policy writers' (system) understandings and appreciation of the classroom (site) reality of testing.

These findings are representative of issues within standardised testing, on a national and an international scale, as identified and confirmed through the literature review. The findings initially appear atomistic due to the artificial or arbitrary boundaries imposed in their presentation in the form of separate chapters in this thesis. However, the research emphasises the existence of strong connections between the five findings, as demonstrated in the literature and data. Similarities and differences in participants' perspectives and responses to the researcher's analysis, observation and questioning contribute to the illumination of a group narrative about standardised literacy testing actors' common experiences. This collective story has been central to the development of recommendations from this research.

7.3 Recommendations

7.3.1 Future research.

One recommendation from this case study is that future research could be broadened in scope and applied across educational systems. The generalisability of this study is potentially constrained due to its small scale and, as such, future research could benefit from expansion of the methodology employed here. A larger sample or different school sites, for instance, could contribute to the findings of this research and expand the field of knowledge.

Different methodological approaches could offer advantages to future research. For example, this study was conducted over the period of one round of tests and concluded following the return to the school of the results of the observed tests; the entire data collection and generation took place over only one school year. A longitudinal study extended for a lengthier period of time

could address broader themes and issues, such as different perspectives on test administration through teachers' careers and the effects upon test actors and test data of changes over time in testing policy and practice.

Further research in the form of replication studies could highlight sources of commonality and divergence in experiences of testing, giving voice to additional site-level realities of testing not raised in this case study by these participants. Such studies could broaden researchers', educators' and policy actors' awareness of effects of standardised literacy testing on classroom test actors in other locations and contexts. A number of interesting questions could be raised as a result of this research. For example, further research could examine the relationship or balance between teachers' maintenance of test rules and their desire to maintain a positive classroom environment. Participant opinions of the inappropriateness or invalidity of such tests for particular students could also form the basis of replication studies, to identify how widespread or legitimate teacher concerns might be about students' need for support during testing. Possible consequences of such assistance upon the value and validity of testing could be a further focus of research.

Further studies might also benefit from research methods which highlight and capture otherwise invisible areas of interest or importance; as an example, an autoethnographic approach within a multi-researcher study could allow for an identification of issues pertinent to the different test actors. This investigation has focused on teachers: teachers' methods of test administration, perspectives of their roles within testing, and interpretations of test value and validity. Future research could direct attention specifically to students,

principals, teachers' aides or parents. The experiences and perspectives of these groups in relation to standardised literacy testing could build on this research for a more comprehensive and nuanced understanding of the total context within which testing is conceived, planned, conducted and developed.

This study has generated and presented case study data of teachers' behaviours, decisions, perspectives and motivations in relation to standardised literacy testing and the test situation. Further research could examine teachers' and test actors' responses to these same or other issues within tests and testing processes. One finding of this research was the participants' problematising of what they saw as an increasing teacher deprofessionalisation and lack of teacher involvement in test construction. Further research could examine the viability of more active involvement of practising teachers in such aspects of standardised literacy testing as test content, policy, timeframes and the conditions under which students are tested.

7.3.2 Policy and professional practice.

A number of the findings of this investigation have clear consequences for education policy, and open up further areas for consideration and debate.

These data particularly make the case for a more nuanced approach to standardised literacy testing policy and practice, and for greater appreciation within policy and practice of variations in classrooms and school contexts. In particular, future research could explore how policy makers might better respond to perceptions identified in this research and allow for reported variations and exceptions to the 'one-size fits all' approach within standardised

literacy testing that was of concern to these participants. ‘Value-added’ test reporting is one suggested way of doing this (Jensen, 2010); future research could examine the efficacy of such test reporting for improving accessibility, relevance and value of test data.

The participant-perceived gulf between system and site perceptions and expectations of standardised literacy testing merits response from test and policy developers. This study recommends that participant cynicism about the uses of student and school literacy data should be accorded serious attention by governments and the system level when instituting literacy testing and when reporting and acting on test data. However, what could also be utilised and built upon is the clear instructional benefits of such testing (Phelps, 2005; Popham, 2001).

A further finding with implications for education concerns participants’ perceptions of a lack of professional respect and appreciation given to their teaching expertise. The data suggest that recognition of teachers’ expertise and opinions in the drafting and implementation of education policy and practice could facilitate a positive educational outcome for teachers within standardised literacy testing (Gardner, 2006). Increased involvement by practising teachers could be considered at various stages of standardised literacy testing, from education policy relating to targets, funding and penalties, to educational construction of tests and to the establishment of reviewed literacy standards.

Participants’ criticism and scepticism of perceived misuse of test data highlight that such data could instead provide governments, testing officials, education departments and faculties with valuable material for policy and

practice including teacher training and professional development. Further, the finding that these participants considered some forms of standardised literacy testing to be inappropriate for their students points to potential benefits in problem-solving dialogue between educators and governments, as well as professional development for teachers in the form of focused training and workshops to address issues in practice. Therefore, this research recommends that educational administrators and policymakers recognise this perceived discord and these system-site tensions within testing as a matter requiring serious and collective attention.

The finding that participants differently interpret test administration rules and guidelines, according to perceptions of student needs, predictions of behavioural issues and intended professional use of test data, has particularly clear and significant policy implications. The effect of such teacher reflexivity within testing upon student literacy standards and data has considerable merit as a future research topic. Educators and students alike would clearly gain from more policy (system) clarity regarding flexibility and reflexivity in school site testing, and consideration of such pressing matters as teacher interpretation and adaptation of test rules and guidelines for their student groups.

7.4 Research Contribution

This research, however small in scale, has made an important contribution to existing knowledge about high stakes, standardised literacy testing with an examination of the ‘lived’ school experience of teachers. This unique examination of teachers in a school site, standardised literacy testing situation

builds onto existing knowledge by showing how these test actors critique, cope and strive to do their best for themselves and their students in response to tensions between expectations and realities.

The tensions created for test actors by the dominant discourses of system demands for uniformity and site realities of reflexivity formed the basis of participants' perceptions and performances. The school site location contextualised the data and provided rich material from which the researcher could focus on particular research questions intended to encapsulate current issues, contestations and dilemmas for test actors, in a context in which "much education policy has been unsuccessful in improving student outcomes or in reducing the inequalities in those outcomes while also having negative effects on educators' morale" (Levin, 2010, p. 739). A series of suggestions and recommendations have been drawn from the findings, and possible future policy, practice and research directions are encouraged.

The findings identified areas of discord and mistranslation between system demands and site realities, which correlated with teachers' perceptions of test value and validity and influenced their approaches to test administration. The findings confirmed that participants saw themselves as often constrained at the site level by power structures stemming from the system level. The overarching system-site tensions throughout the data meant that school site test actors were charged with following particular rules of behaviour, set by those actors located outside the classroom site, elsewhere in the education and political system, and at more senior hierarchical levels. In response to expectations communicated through the machinery of structural power,

participants employed both compliant and noncompliant behaviours that showed their recognition and appreciation of their own positionality.

The research recognised that participants were aware of particular expectations on them as professionals and were concerned about teacher deprofessionalisation as one consequence of test-related education policy. Churchill, Williamson and Grady (1997) and Sim (2006) located moves to return to basics in education and what the public, media and government can easily understand within a general Western uncertainty about what the future holds. Research into literacy developments, however, identified that future literacy needs will be expanded and different from those of the past and present-day, and that “an enriched NAPLAN [would incorporate] more useful assessment of students’ control of multiple modes and channels of communication” (Wyatt-Smith, 2008, p. 61).

Future research could build knowledge of this topic by further exploration of the nuances involved and complexities of teachers’ perspectives and practices regarding standardised literacy testing and test administration identified and examined here. This research argues the importance of greater consideration of the problematic of tensions experienced by teachers with standardised literacy testing. Education policy development needs to take account of the consequences of the contested nature of these ‘high stakes’ tests at the classroom level. This research indicates that by inadequately addressing the complexities and contradictions for teachers involved with standardised literacy testing and test administration, the potential benefits are not being realised. Further, “the purposes that are served and the impact of the testing

and reporting regimes” (Caldwell, 2010, p. 5) need to be more carefully communicated and scrutinised, in order to address these broad issues. Rather than continuing to divide standardised literacy testing and related assessments between what is helpful at the classroom, site level and what is needed by governments and public at the broader system level, Wyatt-Smith (1998) cogently summarised and advised,

We need to get the balance right, keeping the two [testing] goals aligned in the best interests of students, parents and the wider community. To this end, we need to establish a platform of informing propositions about the nature of literacy assessment, as a basis for re-examining what standardised testing can (and cannot) achieve. (p. 90)

Positives and negatives of high stakes, standardised literacy testing can be readily identified and convincingly argued; such polemic arguments clearly do not address the lived, school site realities of testing, but divide teachers, parents, researchers, government and the public. Tankersley (2007) reminded that it is unrealistic and pointless to argue that accountability testing be halted, as such forms of assessment are a very real and apparently permanent fixture of education systems, and that instead teachers should be better informed, prepared and tests should be better designed and used (also Phelps, 2005). Policy development that takes full account of teacher concerns and involves them in problem solving to address outstanding issues is recommended, and at the school site level professional and practice development could address teacher preparation and implementation.

References

- Abbarno, G. (1998). Corporate sponsored education: The limits of social responsibility. Presented at Paideia Project On-Line: Twentieth World Congress of Philosophy, Boston, MA. [Transcript]. Retrieved from <http://www.bu.edu/wcp/Papers/Educ/EducAbba.htm>
- Abrams, L., Pedulia, J., & Madaus, G. (2003). Views from the classroom: Teachers' opinions of statewide testing programs. *Theory into Practice*, 42(1), 18-29.
- Achinstein, B., Ogawa, R., & Speigman, A. (2004). Are we creating separate and unequal tracks of teachers? The effects of state policy, local conditions, and teacher characteristics on new teacher socialization. *American Educational Research Journal*, 41(3), 557-603.
- Afflerbach, P. (2009a). Accountability testing: Getting situated. *Educational Researcher*, 38(6), 468-471.
- Afflerbach, P. (2009b). Response to "The past of test-based educational accountability", by Gregory J. Cizek. *Educational Researcher*, 38(6), 472-473.
- Ainley, P. (2004). The new "market-state" and education. *Journal of Education Policy*, 19(4), 497-514.
- Akiba, M., LeTendre, G., & Scribner, J. P. (2007). Teacher quality, opportunity gap, and national achievement in 46 countries. *Educational Researcher*, 36(7), 369-387.
- Albert, D. (2005). [Introduction]. In J. T. Gatto, *Dumbing us down: The hidden curriculum of compulsory curriculum* (2nd ed., pp. xv-xxvii). Gabriola Island, Canada: New Society Publishers.
- Alloway, N., & Gilbert, P. (1998). Reading literacy test data: Benchmarking success? *The Australian Journal of Language and Literacy*, 21(3), 249-261.
- Alvermann, D. E., & Reinking, D. (2007). Multiple dimensions of achievement: Defining, identifying, and addressing "the gap". *Reading Research Quarterly*, 42(3), 406-428.
- Ambach, G. (1999). Comments on "Education standards". In F. M. Ottobre (Ed.), *The role of measurement and evaluation in education policy* (pp. 49-53). Paris, France: UNESCO.
- Amrein-Beardsley, A. (2008). Methodological concerns about the Education Value-Added Assessment System. *Educational Researcher*, 37(2), 65-75.
- Andersen, S. C. (2008). The impact of public management reforms on student performance in Danish schools. *Public Administration*, 86(2), 541-558.

- Anderson, G. (with Arsenault, N.). (2004). *Fundamentals of educational research* (2nd ed.). London, UK, & New York, NY: Routledge Falmer.
- Anderson, S., Medrich, E., & Fowler, D. (2007). Which achievement gap? *The Phi Delta Kappan*, 88(7), 547-550.
- Apple, M. (1991). [Series editor's introduction]. In P. Lather (Ed.), *Getting smart: Feminist research and pedagogy with/in the postmodern* (pp. vii-xi). New York, NY: Routledge.
- Apple, M. (1993). *Official knowledge: Democratic education in a conservative age*. New York, NY: Routledge.
- Apple, M. (1996). *Cultural politics and education*. Buckingham, UK: Open University Press.
- Apple, M. (2001). *Educating the "right" way: Markets, standards, God, and inequality*. New York, NY: Routledge Falmer.
- Apple, M., & Beane, J. (2007). Lessons from democratic schools. In J. Beane & M. Apple (Eds.), *Democratic schools: Lessons in powerful education* (2nd ed., pp. 150-155). Portsmouth, N.H.: Heinemann.
- Applebaum, B. (2001). Locating oneself, self "T"-dentification and the trouble with moral agency. *Philosophy of Education*, 412-422. Retrieved from <http://www.ed.uiuc.edu/eps/PES-Yearbook/2001/applebaum%2001.pdf>
- Atkinson, P. (2004). Performance, culture and the sociology of education. *International Studies in the Sociology of Education*, 14(2), 147-165.
- Au, W. (2007). High-stakes testing and curricular control: A qualitative metasynthesis. *Educational Researcher*, 36(5), 258-267.
- Au, W. (2008). Teaching in dystopia. *Rethinking Schools*, 22(3). Retrieved from <http://www.rethinkingschools.org>
- Australian Curriculum, Assessment and Reporting Authority (ACARA) (2010a). *National Assessment Program – Literacy and Numeracy: Home*. Retrieved from <http://www.naplan.edu.au/>
- Australian Curriculum, Assessment and Reporting Authority (ACARA) (2010b). *National Assessment Program – Literacy and Numeracy: National Report*. Retrieved from http://www.naplan.edu.au/reports/national_report.html
- Australian Curriculum, Assessment and Reporting Authority (ACARA) (2010c). *National Assessment Program – Literacy and Numeracy: National summary report: Achievement in reading, writing, language conventions and numeracy 2010*. Retrieved from http://www.naplan.edu.au/verve/resources/NAPLAN_2010_Summary_Report.pdf
- Bagley, C. (2006). School choice and competition: A public-market in education revisited. *Oxford Review of Education*, 32(3), 347-362.

- Baines, L. (2007). Learning from the world: Achieving more by doing less. *The Phi Delta Kappan*, 89(2), 98-100.
- Baker, K. (2007). Are international tests worth anything? *The Phi Delta Kappan*, 89(2), 101-104.
- Ball, D. L., & Forzani, F. (2007). What makes education research “educational”? *Educational Researcher*, 36(9), 529-540.
- Ball, S. J. (2009). Privatising education, privatising education policy, privatising educational research: Network governance and the “competition state”. *Journal of Education Policy*, 24(1), 83-99.
- Ballis, C. M. (1999/2000). Meeting the challenge of literacy for all. *English in Australia*, 126, 30-42.
- Bamford, P. (2010, April 7). Teaching to the test. *Inside Story*. Retrieved from <http://www.inside.org.au/teaching-to-the-test>
- Blanchard, J. (2003). Targets, assessment for learning, and whole-school improvement. *Cambridge Journal of Education*, 33(2), 257-271.
- Beane, J., & Apple, M. (2007). The case for democratic schools. In J. Beane & M. Apple (Eds.), *Democratic schools: Lessons in powerful education* (2nd ed., pp. 1-29). Portsmouth, NH: Heinemann.
- Beare, H. (2006). *How we envisage schooling in the 21st century: Applying the new ‘imaginary’*. London, UK: Specialist Schools and Academies Trust. Retrieved from <http://www.ssat-inet.net/resources.aspx>
- Beck, J. (2008). Government professionalism: Re-professionalising or de-professionalising teachers in England? *British Journal of Educational Studies*, 56(2), 119-143.
- Beck, J., & Young, M. (2005). The assault on the professions and the restructuring of academic and professional identities: A Bernsteinian analysis. *British Journal of Education Studies*, 26(2), 183-197.
- Benjamin, S. (2003). What counts as “success”? Hierarchical discourses in a girl’s comprehensive school. *Discourse: Studies in the Cultural Politics of Education*, 24(1), 105-118.
- Berk, R. A. (2002). *Humor as an instructional defibrillator: Evidence-based techniques in teaching and assessment*. Sterling, VA: Stylus.
- Berliner, D. C. (2001). Learning about and learning from experts. *International Journal of Educational Research*, 35(5), 463-482.
- Berliner, D. C. (2002). Educational research: The hardest science of all. *Educational Researcher*, 31(8), 18-20.

- Berliner, D., & Biddle, B. (1995). *The manufactured crisis: Myths, fraud, and the attack on America's public schools*. Reading, MA: Perseus Books.
- Bernal, J. L. (2005). Parental choice, social class and market forces: The consequences of privatization of public services in education. *Journal of Education Policy*, 20(6), 779-792.
- Beutel, D. (2010). The nature of pedagogic teacher-student interactions: A phenomenographic study. *The Australian Educational Researcher*, 37(2), 77-91.
- Black, S. (2004). Whose economic wellbeing? A challenge to dominant discourse on the relationship between literacy/numeracy skills and (un)employment. *Literacy & Numeracy Studies*, 13(1), 7-17.
- Black, W. R. (2004). *"Equal access to mandated testing": Policies, disciplinary discourse, and practices of performance in the lives of English language learner youth* (Unpublished doctoral dissertation). The University of Texas at Austin, Austin, TX.
- Blaxter, L., Hughes, C., & Tight, M. (1996). *How to Research*. Buckingham, UK: Open University.
- Bloome, D., & Clark, C. (2006). Discourse-in-use. In J. Green, G. Camilli, & P. Elmore (Eds.), *Handbook of complementary methods in education research* (pp. 227-241). Mahwah, NJ: Lawrence Erlbaum.
- Blum, R. (2000). Standards-based reform: Can it make a difference for students? *Peabody Journal of Education*, 75(4), 90-113.
- Booher-Jennings, J. (2005). Below the bubble: "Educational triage" and the Texas Accountability System. *American Educational Research Journal*, 42(2), 231-268.
- Boomer, G. (1998). English teaching: Art & science. *English in Australia*, 122, 15-25.
- Bracey, G. (2007). Reports, reports. *The Phi Delta Kappan*, 88(9), 718-720.
- Bracey, G. (2008). Competing concepts of competitiveness. *The Phi Delta Kappan*, 89(8), 620-621.
- Braun, H., Zhang, J., & Vezzu, S. (2010). An investigation of bias in reports of the National Assessment of Educational Progress. *Educational Evaluation and Policy Analysis*, 32(1), 24-43.
- Brenner, M. (2006). Interviewing in educational research. In J. Green, G. Camilli, & P. Elmore (Eds.), *Handbook of complementary methods in education research* (pp. 357-370). Mahwah, N.J.: Lawrence Erlbaum.
- Brewer, D., & Goldhaber, D. (2008). Examining the incentives in educational research. *The Phi Delta Kappan*, 89(5), 361-364.

- Britzman, D. P. (1995) Teachers under suspicion: Is it true that teachers aren't as good as they used to be? In J. L. Kincheloe & S. R. Steinberg (Eds.), *Thirteen questions: Reframing education's conversation* (2nd ed., pp. 69-76). New York, NY: Peter Lang.
- Brookhart, S., & Bronowicz, D. (2003). "I don't like writing. It makes my fingers hurt": Students talk about their classroom assessments. *Assessment in Education: Principles, Policy and Practice*, 10(2), 221-242.
- Brooks, J., Maxcy, B., & Nguyen, T. (2010). International perspectives on the politics of education: Leadership & identity in multiple contexts: An introduction to the Politics of Education special issue [Introduction]. *Educational Policy*, 24(1), 3-8.
- Brown, C. (2008). Keep it cheap, keep it local, and keep it coming: Standards-based accountability reform in Wisconsin. *Educational Policy*, 22(2), 250-294.
- Brown, G. (2004). Teachers' conceptions of assessment: Implications for policy and professional development. *Assessment in Education*, 11(3), 301-318.
- Brown, K. D. (2005). C'mon, tell me. ... Does school ethnography really matter? *Educational Researcher*, 34(9), 29-34.
- Brown, M., McCallum, B., Taggart, B., & Gipps, C. (1997). The validity of national testing at age 11: The teacher's view. *Assessment in Education*, 4(2), 271-293.
- Buckingham, J. (2003, January 2). Parents need report cards on schools. *The Sydney Morning Herald*. Retrieved from <http://www.cis.org.au>
- Budd, Y. (2009). *Any bodies out there?: Educating the posthuman*. (Unpublished doctoral dissertation). University of Tasmania, Hobart, TAS.
- Bulterman-Bos, J. (2008). Will a clinical approach make education research more relevant for practice? *Educational Researcher*, 37(7), 412-420.
- Burch, P. (2010). After the fall: Educational contracting in the USA and the global financial crisis. *Journal of Education Policy*, 25(6), 757-766.
- Burns, R. (2000). *Introduction to research methods* (4th ed.). Frenchs Forest, NSW: Pearson Education.
- Busher, H. (2006). Coping with policy in social contexts in schools [Editorial]. *Improving Schools*, 9(2), 91-92.
- Butler, T., & van Zanten, A. (2007). School choice: A European perspective [Editorial]. *Journal of Education Policy*, 22(1), 1-5.
- Caldwell, B. (2010). *The impact of high stakes test driven accountability*. Retrieved from <http://www.aeufederal.org.au/>
- Carabine, J. (2001). Unmarried motherhood 1830-1990: A genealogical analysis. In M. Wetherell, S. Taylor, & S. J. Yates (Eds.), *Discourse as data: A guide for analysis* (pp. 267-307). London, UK & Thousand Oaks, CA: Sage.

- Carroll, T. (2002). Writing my way to poststructuralism. *Outskirts: Feminisms Along the Edge*, 9. Retrieved from <http://www.chloe.uwa.edu.au/outskirts/archive/VOL9/article3.html>
- Castleton, G. (2010, 9 September). *What can we learn from other countries that will inform how community groups engage with adults with limited literacy skills?* Lecture presented for the Tasmanian Community Fund conference, Launceston, TAS. [Powerpoint]. Retrieved from <http://www.tascomfund.org/>
- Castleton, G., Wyatt-Smith, C., Cooksey, R., & Freebody, P. (2003). The nature of teachers' qualitative judgements: A matter of context and salience. Part two: One-of-context judgements. *Australian Journal of Language and Literacy*, 26(2), 33-42.
- Charmaz, K. (2000). Grounded theory: Objectivist and constructivist methods. In N.K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 509-535). Thousand Oaks, CA: Sage.
- Charmaz, K. (2001). Qualitative interviewing and grounded theory analysis. In J. F. Gubrium & J. A. Holstein (Eds.), *Handbook of interview research: Context & method* (pp. 675-694). Thousand Oaks, CA: Sage.
- Charmaz, K. (2005). Grounded theory in the 21st century. In N. K. Denzin & Y. S. Lincoln (Eds.), *The Sage handbook of qualitative research* (3rd ed., pp. 507-535). Thousand Oaks, CA: Sage.
- Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative analysis*. London, UK & Thousand Oaks, CA: Sage.
- Charmaz, K., & Mitchell, R. (2002). Grounded theory in ethnography. In P. Atkinson, A. Coffey, S. Delamont, J. Lofland, & L. Lofland (Eds.), *The handbook of ethnography* (pp. 160-174). London, UK: Sage.
- Christensen, L. (2008). The power of words. *Rethinking Schools*, 22(3). Retrieved from <http://www.rethinkingschools.org>
- Churchill, R., Williamson, J., & Grady, N. (1997). Educational change and the new realities of teachers' work lives. *Asia-Pacific Journal of Teacher Education*, 25(2), 141-158.
- Cizek, G. (2009a). Assessment for accountability in education: Past, present, and challenges ahead. *Educational Researcher*, 38(6), 467-473.
- Cizek, G. (2009b). Response to "Accountability Testing: Getting Situated", by Peter P. Afflerbach. *Educational Researcher*, 38(6), 467-473.
- Clarke, S. (2001). Reforming teachers' work and changing roles for unions. *Education Research and Perspectives*, 28(2), 33-50.
- Clegg, S. (2005). Evidence-based practice in educational research: A critical realist critique of systematic review. *British Journal of Sociology of Education*, 26(3), 415-428.

- Cohen, J. (2010). Teachers in the news: A critical analysis of one US newspaper's discourse on education, 2006-2007. *Discourse: Studies in the Cultural Politics of Education*, 31(1), 105-119.
- Cohen, L., Manion, L., & Morrison, K. (2000). *Research methods in education* (5th ed.). London, UK: Routledge Falmer.
- Coldron, J., & Smith, R. (1999). The construction of reflective practice in key policy documents in England. *Pedagogy, Culture & Society*, 7(2), 305-320.
- Cole, P. (2010). *From "merit pay" to "pay for contribution": New developments in teacher performance pay*. Centre for Strategic Education Seminar Series Paper No. 191. Retrieved from <http://www.apo.org.au/>
- Coleman, S. (2008). Fault lines in merit pay. *Rethinking Schools*, 22(4). Retrieved from <http://www.rethinkingschools.org>
- Collidge, L. (2001). Negotiating the curriculum. *English in Australia*, 132, 76-78.
- Collin, R., & Apple, M. (2007). Schooling, literacies and biopolitics in the Global Age. *Discourse: Studies in the Cultural Politics of Education*, 28(4), 433-454.
- Comber, B., & Kamler, B. (2004). Getting out of deficit: Pedagogies of reconnection. *Teaching Education*, 15(3), 293-310.
- Comber, B., & Nixon, H. (2009). Teachers' work and pedagogy in an era of accountability. *Discourse: Studies in the Cultural Politics of Education*, 30(3), 333-345.
- Cooksey, R., Freebody, P., & Wyatt-Smith, C. (2007). Assessment as judgement-in-context: Analysing how teachers evaluate students' writing. *Educational Research and Evaluation*, 13(5), 401-434.
- Corbin, J., & Strauss, A. (1990). Grounded theory research: Procedures, canons, and evaluative criteria. *Qualitative Sociology*, 13(1), 3-21.
- Cotton, K. (1995). *Effective schooling practices: A research synthesis*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Craig, N. (2010, September 5). Literacy tests "full of flaws". *The Age*. Retrieved from <http://www.theage.com.au/>
- Crow, G., & Weindling, D. (2010). Learning to be political: New English headteachers' roles. *Educational Policy*, 24(1), 137-158.
- Cumming, J. J., Wyatt-Smith, C., Elkins, J., & Neville, M. (2006). *Teacher judgment: Building an evidentiary base for quality literacy and numeracy education*. Griffith University, NSW.
- Daniels, B. (2005, 27-29 July). *Independent schools and accountability: The national agenda*. Lecture presented at AISQ 2005 Biennial State Conference, Townsville, QLD. [Transcript]. Retrieved from <http://www.isca.edu.au/>

- Davies, B., & Bansel, P. (2007). Neoliberalism and education. *International Journal of Qualitative Studies in Education*, 20(3), 247-259.
- DeBray-Pelot, E., & McGuinn, P. (2009). The new politics of education: Analyzing the Federal education policy landscape in the post-NCLB era. *Educational Policy*, 23(1), 15-42.
- DeCoker, G. (2002a). What do national standards really mean? [Introduction]. In G. DeCoker (Ed.), *National standards and school reform in Japan and the United States* (pp. xi-xx). New York, NY: Teachers College Press, Columbia University.
- DeCoker, G. (2002b). The creation of national standards: Influences from above and below. In G. DeCoker (Ed.), *National standards and school reform in Japan and the United States* (pp. 1-3). New York, NY: Teachers College Press, Columbia University.
- Delamont, S. (1976). *Interaction in the classroom*. London, UK: Methuen.
- Demaine, J. (2005). The “new” politics of education in Britain’s changing times. *International Studies in Sociology of Education*, 15(2), 115-127.
- Den Brok, P., Levy, J., Brekelmans, M., & Wubbels, T. (2005). The effect of teacher interpersonal behaviour on students’ subject-specific motivation. *The Journal of Classroom Interaction*, 40(2), 20-33.
- Department of Education (Tasmania). (2010) *School levy policy*. Retrieved from <http://www.education.tas.gov.au/school/parents/costs/levy>
- Dey, I. (1999). *Grounding theory: Guidelines for qualitative inquiry*. San Diego, CA: Academic Press.
- Dey, I. (2007). Grounding categories. In A. Bryant & K. Charmaz (Eds.), *The SAGE handbook of grounded theory* (pp. 167-190). London, UK: Sage.
- Doecke B. (2007). Beyond externalising and finalising definitions: Standards for Teachers of English Language and Literacy in Australia (STELLA). *English in Australia*, 42(3), 31-42.
- Doecke, B., Reynolds, G., & Roberts, A. (2002). Standardised testing: What space for professional judgement? *English in Australia*, 135, 5-8.
- Donnelly, K. (2004). *Why our schools are failing*. Sydney, NSW: Duffy & Snellgrove.
- Dressman, M. (2007). Theoretically framed: Argument and desire in the production of general knowledge about literacy. *Reading Research Quarterly*, 42(3), 332-363.
- Duffield, J., Allan, J., Turner, E., & Morris, B. (2000). Pupils’ voices on assessment: An alternative to the standards agenda. *Cambridge Journal of Education*, 30(2), 263-274.

- Dugdale, G., & Clark, C. (2008). *Literacy changes lives: An advocacy resource*. London, UK: National Literacy Trust. Retrieved from http://www.literacytrust.org.uk/research/nlt_research/243_literacy_changes_lives_an_advocacy_resource
- Dwyer, J. (2004). Introduction to symposium: School accountability and “high stakes” testing. *Theory and Research in Education*, 2(3), 211-217.
- Edley, N. (2001). Analysing masculinity: Interpretative repertoires, ideological dilemmas and subject positions. In M. Wetherell, S. Taylor, & S. Yates (Eds.), *Discourse as data: A guide for analysis* (pp. 189-228). London, UK & Thousand Oaks, CA: Sage.
- EdSource. (2002). *California student achievement: Multiple views of K-12 progress*. Palo Alto, CA: Author.
- Eisenhart, M., & Borko, H. (1993). *Designing classroom research: Themes, issues, and struggles*. Needham Heights, MA: Allyn & Bacon.
- Emmer, E. T., & Gerwels, M. C. (2006). Classroom management in middle and high school classrooms. In C. M. Evertson & C. S. Weinstein (Eds.), *Handbook of classroom management: Research, practice, and contemporary issues* (pp. 407-437). New York, NY: Routledge.
- English, E., Hargreaves, L., & Hislam, J. (2002). Pedagogical dilemmas in the National Literacy Strategy: Primary teachers’ perceptions, reflections and classroom behaviour. *Cambridge Journal of Education*, 32(1), 9-26.
- Ercikan, K., & Roth, W.M. (2006). What good is polarizing research into qualitative and quantitative? *Educational Researcher*, 35(5), 14-23.
- Ertl, H. (2006). Educational standards and the changing discourse on education: The reception and consequences of the PISA study in Germany. *Oxford Review of Education*, 32(5), 619-634.
- Everson, H. T. (1999). Education standards: Current directions and implications for assessment. In F. M. Ottobre (Ed.), *The role of measurement and evaluation in education policy* (pp. 39-49). Paris, France: UNESCO.
- Evertson, C. M., & Weinstein, C. S. (2006). Classroom management as a field of inquiry. In C. M. Evertson & C. S. Weinstein (Eds.), *Handbook of classroom management: Research, practice, and contemporary issues* (pp. 1-15). New York, NY: Routledge.
- Fairclough, N. (2001). The discourse of New Labour: Critical discourse analysis. In M. Wetherell, S. Taylor, & S. Yates (Eds.), *Discourse as data: A guide for analysis* (pp. 229-266). London, UK & Thousand Oaks, CA: Sage.
- FairTest. (2008). Standardized testing has serious limitations. In D. A. Henningfield (Ed.), *Standardized testing* (pp. 17-24). Detroit, MI: Thomson Gale.

- Fehring, H. (2005). Critical, analytical and reflective literacy assessment: Reconstructive practice. *Australian Journal of Language and Literacy*, 28(2), 95-113.
- Felson, M. (1982). The web of learning. In K. J. Gilmartin & R. J. Rossi (Eds.), *Monitoring educational outcomes and public attitudes* (pp. 37-51). New York, NY: Human Sciences Press.
- Ferrari, J. (2007, May 11). Standardised tests fail students, say teachers. *The Australian*. Retrieved from <http://www.theaustralian.com.au/>
- Finnigan, K., & Gross, B. (2007). Do accountability policy sanctions influence teacher motivation? Lessons from Chicago's low-performing schools. *American Educational Research Journal*, 44(3), 594-629.
- Fiske, E., & Ladd, H. (2000). *When schools compete: A cautionary tale*. Washington, DC: Brookings Institution Press.
- Florian, L., & Rouse, M. (2001). Inclusive practice in English secondary schools: Lessons learned. *Cambridge Journal of Education*, 31(3), 399-412.
- Foucault, M. (1980). *Power/knowledge: Selected interviews and other writings, 1972-1977* (C. Gordon, Ed.). New York, NY: Pantheon.
- Freebody, P. (1998). Findings from the literacy scandal: "The same disconcert and tremulousness and meditation". *English in Australia*, 122, 10-14.
- Freebody, P. (2004). *Qualitative research in education: Interaction and practice*. London, UK: Sage.
- Freebody, P. (2007). *Literacy education in school: Research perspectives from the past, for the future*. Camberwell, VIC: Australian Council for Educational Research (ACER) Press.
- Freebody, P., & Wyatt-Smith, C. (2004). The assessment of literacy: Working the zone between "system" and "site" validity. *Journal of Educational Enquiry*, 5(2), 30-49.
- Freeman, C. (2009). First national literacy and numeracy tests introduced. *Research Developments*, 20(20), 12-15. Retrieved from <http://research.acer.edu.au/resdev/vol20/iss20/12/>
- Fremer, J. (2005). Foreword: The rest of the story [Foreword]. In R. Phelps (Ed.), *Defending standardized testing*. Mahwah, NJ: Lawrence Erlbaum.
- French, D. (2008). High-stakes testing is flawed. In D. A. Henningfield (Ed.), *Standardized testing* (pp. 61-72). Detroit, MI: Thomson Gale.
- Friedman, I. A. (2006). Classroom management and teacher stress and burnout. In C. M. Evertson & C. S. Weinstein (Eds.), *Handbook of classroom management: Research, practice, and contemporary issues* (pp. 925-944). New York, NY: Routledge.

- Fuhrman, S. (2002). [Foreword]. In G. DeCoker (Ed.), *National standards and school reform in Japan and the United States* (pp. vii-viii). New York, NY: Teachers College Press, Columbia University.
- Furlong, V. (1976). Interaction sets in the classroom: Towards a study of pupil knowledge. In M. Stubbs & S. Delamont (Eds.), *Explorations in classroom observation* (pp. 23-44). New York, NY: John Wiley & Sons.
- Fusarelli, L. D. (2008). Flying (partially) blind: school leaders' use of research in decision making. *The Phi Delta Kappan*, 89(5), 365-368.
- Gale, T., & Densmore, K. (2002). Student success and failure: As a matter of fact or just how they are portrayed? *Asia-Pacific Journal of Teacher Education*, 30(1), 7-23.
- Gardner, C. (2006). *An opportunity missed?: Trajectories of one education policy in the Tasmania Government school system*. (Unpublished doctoral dissertation). University of Tasmania, Hobart, TAS.
- Gardner, D. (2007). Confronting the achievement gap. *The Phi Delta Kappan*, 88(7), 542-546.
- Garrett, T. (2008). Student-centered and teacher-centered classroom management: A case study of three elementary teachers. *The Journal of Classroom Interaction*, 43(1), 34-47.
- Gawlik, M. (2009). When accountability comes knocking, how do charter schools respond? *eJournal of Education Policy*. Retrieved from <https://www4.nau.edu/cee/jep/journals.aspx?id=301>
- Gee, J. P. (1999). *An introduction to discourse analysis: Theory and method*. London, UK & New York, NY: Routledge.
- Gee, J. P. (2003). Opportunity to learn: A language-based perspective on assessment. *Assessment in Education: Principles, Policy & Practice*, 10(1), 27-46.
- Gee, J. P. (2005). *An introduction to discourse analysis: Theory and method* (2nd ed.). London, UK, & New York, NY: Routledge.
- Gee, J. P. (2011). *An introduction to discourse analysis: Theory and method* (3rd ed.). London, UK, & New York, NY: Routledge.
- Geertz, C. (1973). *The interpretation of cultures: Selected essays by Clifford Geertz*. New York, NY: Basic Books.
- Gill, M. (1998). Who set the benchmarks? Analysing the national literacy agenda. *English in Australia*, 121, n.p.
- Gillham, W. (2000). *Case study research methods*. London, UK & New York, NY: Continuum.

- Gipps, C. (1999). Equity in education and assessment. In F. M. Ottobre (Ed.), *The role of measurement and evaluation in education policy* (pp. 15-28). Paris, France: UNESCO.
- Giroux, H. (1995). Educational visions: What are schools for and what should we be doing in the name of education? In J. L. Kincheloe & S. R. Steinberg (Eds.), *Thirteen questions: Reframing education's conversation* (2nd ed., pp. 295-303). New York, NY: Peter Lang.
- Goffman, E. (1990). *The presentation of self in everyday life* (2nd ed.). London, UK: Penguin.
- Goodwyn, A. (2000/2001). Who wants to be a super teacher? The perils and pleasures of recognising expertise in English teaching. *English in Australia*, 129-130, 39-50.
- Goodwyn, A. (2004). The professional identity of English teachers. *English in Australia*, 139, 122-130.
- Goodwyn, A., & Findlay, K. (2002). Language, literacy and the discourses of English teaching in England: A case study. *Educational Studies in Language and Literature*, 2, 221-238.
- Gorard, S. (2006). Value-added is of little value. *Journal of Educational Policy*, 21(2), 235-243.
- Gorard, S., & Fitz, J. (2006). What counts as evidence in the school choice debate? *British Educational Research Journal*, 32(6), 797-816.
- Gordon, J. E. (1999). Comments on "Equity in education and assessment". In F. M. Ottobre (Ed.), *The role of measurement and evaluation in education policy* (pp. 28-31). Paris, France: UNESCO.
- Gordon, L., & Whitty, G. (1997). Giving the "Hidden Hand" a helping hand? The rhetoric and reality of neoliberal education reform in England and New Zealand. *Comparative Education*, 33(3), 453-467.
- Grant, N. (2009). Schools of little thought: Why change management hasn't worked. *Improving Schools*, 12(1), 19-32.
- Green, B. (2006). English, literacy, rhetoric: Changing the project? *English in Education*, 40(1), 7-19.
- Green, B., Hodgins, J., & Luke, A. (1997). Debating literacy in Australia: History lessons and popular f(r)ictions. *The Australian Journal of Language and Literacy*, 20(1), 6-24.
- Greene, R. (2008). Kids do well if they can. *The Phi Delta Kappan*, 90(3), 160-167.
- Haig, B. (1995). Grounded theory as scientific method. *Philosophy of Education*. Retrieved from http://www.ed.uiuc.edu/EPS/PES-yearbook/95_docs/haig.html

- Hanushek, E. A. (2001). Incentives: Linking resources, performance, and accountability. In *The Jossey-Bass reader on school reform* (pp. 299-334). San Francisco, CA: Jossey-Bass.
- Hargreaves, D. (2004). *Personalising learning, series 2: Student voice and assessment for learning*. London, UK: Specialist Schools and Academies Trust. Retrieved from <http://www.ssat-inet.net/resources.aspx>
- Harlen, W. (2005). Trusting teachers' judgement: Research evidence of the reliability and validity of teachers' assessment used for summative purposes. *Research Papers in Education*, 20(3), 245-270.
- Harris, S. (2005). Professionals, partnerships and learning in changing times. *International Studies in Sociology of Education*, 15(1), 71-86.
- Hatch, J. (2002). *Doing qualitative research in education settings*. Albany, NY: State University of New York Press.
- Hawkins, B. R. (2008). Professional development. In S. B. Kucer (Ed.), *What research REALLY says about teaching and learning to read* (pp. 250-271). Urbana, IL: National Council of Teachers of English.
- Hemsley-Brown, J. (2004). Facilitating research utilisation: A cross-sector review of research evidence. *The International Journal of Public Sector Management*, 17(6), 534-552.
- Henderson, D. (2005). What is education for? Situating history, cultural understandings and studies of society and environment against neo-conservative critiques of curriculum reform. *Australian Journal of Education*, 49(3), 306-319.
- Henig, J. (2009). Politicization of evidence: Lessons for an informed democracy. *Educational Policy*, 23(1), 137-160.
- Henningfield, D. A. (2008). [Introduction]. In D. A. Henningfield (Ed.), *Standardized testing* (pp. 7-9). Detroit, MI: Thomson Gale.
- Hill, D. (2005). Globalisation and its educational discontents: Neoliberalisation and its impacts on education workers' rights, pay and conditions. *International Studies in Sociology of Education*, 15(3), 257-288.
- Hilton, M. (2006). Measuring standards in primary English: Issues of validity and accountability with respect to PIRLS and National Curriculum test scores. *British Educational Research Journal*, 32(6), 817-837.
- Hirsch, E. D. (2007). The tests we need. In Education Week, *The last word: The best commentary and controversy in American education* (pp. 70-78). San Francisco, CA: John Wiley & Sons, Jossey-Bass.
- Ho, I. (2004). A comparison of Australian and Chinese teachers' attributions for student problem behaviors. *Educational Psychology*, 24(3), 375-391.

- Hodgetts, K. (2010). Boys' underachievement and the management of teacher accountability. *Discourse: Studies in the Cultural Politics of Education*, 31(1), 29-43.
- Honan, E. (2010). Mapping discourses in teachers' talk about using digital texts in classrooms. *Discourse: Studies in the Cultural Politics of Education*, 31(2), 179-193.
- Hostetler, K. (2005). What is "good" education research? *Educational Researcher*, 34(6), 16-21.
- Hoy, A. W., & Weinstein, C. S. (2006). Student and teacher perspectives on classroom management. In C. M. Evertson & C. S. Weinstein (Eds.), *Handbook of classroom management: Research, practice, and contemporary issues* (pp. 181-219). New York, NY: Routledge.
- Hudson, P. (2010, February 1). Teacher scorecards. *Herald Sun*. Retrieved from <http://www.heraldsun.com.au/>
- Hudson, P., & Masanauskas, J. (2010, May 6). Union drops planned ban on literacy and numeracy tests. *Herald Sun*. Retrieved from <http://www.heraldsun.com.au/>
- Humphries, E., Mertens, D., & Truman, C. (2000). Arguments for an "emancipatory" research paradigm. In C. Truman, D. Mertens, & E. Humphries (Eds.), *Research and inequality* (pp. 3-23). London, UK: University College of London Press.
- Hursh, D. (2007). Assessing No Child Left Behind and the rise of neoliberal education policies. *American Educational Research Journal*, 3, 493-518.
- Hursh, D. (2008). *High-stakes testing and the decline of teaching and learning: The real crisis in education*. Lanham, MD: Rowman & Littlefield.
- Imel, S., Kerka, S., & Wonacott, M. (2002) Qualitative research in adult, career, and career-technical education. *Practitioner File*, ERIC/ACVE. Retrieved from <http://www-tcall.tamu.edu/erica/docs/pfile05.htm>
- Isaac, S., & Michael, W. B. (1995). *Handbook in research and evaluation*. San Diego, CA: Ed ITS.
- Jensen, B. (2010a). *Measuring what matters: Student progress, Executive summary*. Grattan Institute Report No. 2010-1. Retrieved from http://www.grattan.edu.au/events/event_measuring_what_matters.html
- Jensen, B. (2010b, April 15). NAPLAN tests help champion equality in the classroom. *The Age*. Retrieved from <http://www.theage.com.au/>
- Johnson, D. D., & Johnson, B. (2006). *High stakes: Poverty, testing, and failure in American schools* (2nd ed.). Lanham, MD: Rowman & Littlefield.

- Johnson, D. D., Johnson, B., Farenga, S. J., & Ness, D. (2008). *Stop high-stakes testing: An appeal to America's conscience*. Lanham, MD: Rowman & Littlefield.
- Johnson, R. (2010). The child's right to an education: "Consensus-Minus-One"? *International Journal of Children's Rights*, 18, 185–216.
- Johnston, J. (2001). Locating oneself: A plea for more critical awareness. *Philosophy of Education*, 423-425. Retrieved from <http://www.ed.uiuc.edu/eps/PES-Yearbook/2001/johnston%2001.pdf>
- Kelly, A., & Downey, C. (2010). Value-added measures for schools in England: Looking inside the 'black box' of complex metrics. *Educational Assessment, Evaluation and Accountability*, 22(3), 181-198.
- Kelly, G., Luke, A., & Green, J. (2008). What counts as knowledge in educational settings: Disciplinary knowledge, assessment, and curriculum [Introduction]. *Review of Research in Education*, 32, vii–x.
- Kelly, S., & Monczunski, L. (2007). Overcoming the volatility in school-level gain scores: A new approach to identifying value added with cross-sectional data. *Educational Researcher*, 36(5), 279-287.
- Kennedy, M. M. (2008). Sorting out teacher quality. *The Phi Delta Kappan*, 90(1), 59-63.
- Kerin, R., & Comber, B. (2008). Assessing the risks and possibilities: Standardised literacy testing at Year 9. *Curriculum Perspectives*, 28(3), 65-70.
- Kincheloe, J. L. (2003). *Teachers as researchers: Qualitative inquiry as a path to empowerment* (2nd ed.). London, UK: Routledge Falmer.
- Knapp, P. (1998). Raising the hurdles: The educational politics of unrealistic standards. *English in Australia*, 123, 30-35.
- Kohn, A. (2000a). *The schools our children deserve: Moving beyond traditional classrooms and "tougher standards"*. New York, NY: Houghton Mifflin.
- Kohn, A. (2000b). *The case against standardized testing: Raising the scores, ruining the schools*. Portsmouth, NH: Heinemann.
- Koretz, D. (2008). *Measuring up: What educational testing really tells us*. Cambridge, MA: Harvard University Press.
- Kosar, K. (2005). *Failing grades: The Federal politics of education standards*. Boulder, CO: Lynne Rienner.
- Kostogriz, A. (2008). Becoming a teacher in neo-liberal conditions. *Pedagogy, Culture & Society*, 16(3), 321-325.
- Kozol, J. (1992). *Savage inequalities: Children in America's schools* (2nd ed.). New York, NY: Harper Perennial.

- Kozol, J. (2005). *The shame of the nation: The restoration of apartheid schooling in America*. New York, NY: Crown.
- Kulinna, P. (2007/2008). Teachers' attributions and strategies for student misbehavior. *The Journal of Classroom Interaction*, 42(2), 21-30.
- Kumar, R. (2005). *Research methodology: A step-by-step guide for beginners* (2nd ed.). London, UK & Thousand Oaks, CA: Sage.
- Ladwig, J. (2010). *What NAPLAN and MySchool don't address (but could, and should)*. Lecture presented at the AEU, AGPPA, ASPA National Symposium, Sydney, NSW.
- Lagemann, E. C. (2000). *An elusive science: The troubling history of education research*. Chicago, IL: University of Chicago Press.
- Lake Corporate Consulting (2006). *Standardised literacy and numeracy test scores and 'doing' the New Basics*. Retrieved from education.qld.gov.au/corporate/newbasics/pdfs/litnum_rpt.pdf
- Lankshear, C., & Knobel, M. (2004). *A Handbook for Teacher Research: From design to Implementation*. Maidenhead, UK: Open University Press.
- Lather, P. (1991). *Getting smart: Feminist research and pedagogy with/in the postmodern*. New York, NY: Routledge.
- Leading article: We can't ignore the teachers' boycott of tests. (2010, April 17). *The Independent*. Retrieved from <http://www.independent.co.uk/news/education/schools/leading-article-we-cant-ignore-the-teachers-boycott-of-tests-1950137.html>
- Learmonth, M. (2005). Doing things with words: The case of "management" and 'administration'. *Public Administration*, 83(3), 617-637.
- Lee, J. (2010). Trick or treat: new ecology of education accountability system in the USA. *Journal of Education Policy*, 25(1), 73-93.
- Lee, J., & Wong, K. K. (2004). The impact of accountability on racial and socioeconomic equity: Considering both school resources and achievement outcomes. *American Educational Research Journal*, 41(4), 797-832.
- Lesch, L. (2007). *Our results-driven testing culture: How it adversely affects students' personal experience*. Lanham, MD: Rowman & Littlefield Education.
- Levin, B. (2004). Media-government relations in education. *Journal of Education Policy*, 19(3), 271-283.
- Levin, B. (2007). Schools, poverty, and the achievement gap. *The Phi Delta Kappan*, 89(1), 75-76.
- Levin, B. (2010). Governments and education reform: Some lessons from the last 50 years. *Journal of Education Policy*, 25(6), 739-747.

- Levin, H., & Belfield, C. (2003). The marketplace in education. *Review of Research in Education*, 27, 183-219.
- Lewis, C. (2006). At last: "What's discourse got to do with it?" A meditation on critical discourse analysis in literacy research. *Research in the Teaching of English*, 40(3), 373-379.
- Lewis, R. (2006). Classroom discipline in Australia. In C. M. Evertson & C. S. Weinstein (Eds.), *Handbook of classroom management: Research, practice, and contemporary issues* (pp. 1193-1214). New York, NY: Routledge.
- Lingard, B. (2009). Global/national pressures on education systems: The Andrew Bell Public Lectures. *Discourse: Studies in the Cultural Politics of Education*, 30(3), 235-238.
- Lingard, B. (2010). Policy borrowing, policy learning: Testing times in Australian schooling. *Critical Studies in Education*, 51(2), 129-147.
- Lubienski, C., Gulosino, C., & Weitzel, P. (2009). School choice and competitive incentives: Mapping the distribution of educational opportunities across local education markets. *American Journal of Education*, 115(4), 601-647.
- Luijten, A. (1999). Comments on "Education standards". In F. M. Ottobre (Ed.), *The role of measurement and evaluation in education policy* (pp. 53-54). Paris, France: UNESCO.
- Luna, C., & Turner, C. L. (2001). The impact of the MCAS: Teachers talk about high-stakes testing. *The English Journal*, 91(1), 79-81.
- Lynch, K., & Moran, M. (2006). Markets, schools and the convertibility of economic capital: The complex dynamics of class choice. *British Journal of Sociology of Education*, 27(2), 221-235.
- Macedo, D. (1995). Power and education: Who decides the forms schools have taken, and who should decide? In J. L. Kincheloe & S. R. Steinberg (Eds.), *Thirteen questions: Reframing education's conversation* (2nd ed., pp. 43-57). New York, NY: Peter Lang.
- Matoush, M., & Fu, D. (2008). The paradoxical situation created by test-driven schooling for multilingual children. In S. B. Kucer (Ed.), *What research REALLY says about teaching and learning to read* (pp. 176-197). Urbana, IL: National Council of Teachers of English.
- McCaffrey, D. F., & Hamilton, L. S. (2007). *Value-added assessment in practice: Lessons from the Pennsylvania Value-Added Assessment System Pilot Project, Technical Report*. Santa Monica / Arlington / Pittsburgh: RAND Education.
- McGaw, B. (2007, June 1). *Why literacy matters to the community*. Lecture presented at the Tasmanian Literacy Summit, Hobart, TAS. [Powerpoint]. Retrieved from <http://www.education.tas.gov.au/school/curriculum/literacy/literacy-summit/McGawpresentation.pdf>

- Merrett, G. (2006). Higher standards, better schools for all – a critique: Can market forces close the social and achievement gap?® *Improving Schools*, 9(2), 93-97.
- Messick, S. J. (1999). Key issues. In F. M. Ottobre (Ed.), *The role of measurement and evaluation in education policy* (pp. 11-14). Paris, France: UNESCO.
- Miller, A. (2006). Contexts and attributions for difficult behavior in English classrooms. In C. M. Evertson & C. S. Weinstein (Eds.), *Handbook of classroom management: Research, practice, and contemporary issues* (pp. 1093-1120). New York, NY: Routledge.
- Mills, K. (2005). Deconstructing binary oppositions in literacy discourse and pedagogy. *Australian Journal of Language and Literacy*, 28(1), 67-82.
- Mills, K. (2008). Will large-scale assessments raise literacy standards in Australian schools? *Australian Journal of Language and Literacy*, 31(3), 211-225.
- Miner, B. (2009). The debate over differentiated pay: The devil is in the details. *Rethinking Schools*, 24(1). Retrieved from <http://www.rethinkingschools.org>
- Morine-Dersheimer, G. (2006). Classroom management and classroom discourse. In C. M. Evertson & C. S. Weinstein (Eds.), *Handbook of classroom management: Research, practice, and contemporary issues* (pp. 127-156). New York, NY: Routledge.
- Muir, T., Beswick, K., & Williamson, J. (2010). Up, close and personal: Teachers' responses to an individualised professional learning opportunity. *Asia-Pacific Journal of Teacher Education*, 38(2), 129-146.
- Munt, V. (2004). The awful truth: A microhistory of teacher stress at Westwood High. *British Journal of Sociology of Education*, 25(5), 577-591.
- Murphy, J. (2009). Turning around failing schools: Policy insights from the corporate, government and nonprofit sectors. *Educational Policy*, 23(6), 796-830.
- Naidoo, M. (2010, December 16). AEU backs retirement pay. *The Mercury*. Retrieved from <http://www.themercury.com.au>
- Naples, N. (2003). *Feminism and method: Ethnography, discourse analysis and activist research*. New York, NY & London, UK: Routledge.
- National Curriculum Board. (2009). *Shape of the Australian Curriculum: English*. Barton, ACT: Author. Retrieved from www.acara.edu.au/news_media/publications.html
- Nelson, S., McGhee, M., Meno, L., & Slater, C. (2007). Fulfilling the promise of educational accountability. *The Phi Delta Kappan*, 88(9), 702-709.
- Nespor, J. (2006). Finding patterns with field notes. In J. Green, G. Camilli, & P. Elmore (Eds.), *Handbook of complementary methods in education research* (pp. 297-308). Mahwah, NJ: Lawrence Erlbaum.

- Neuendorf, K. A. (2002). *The content analysis guidebook*. Thousand Oaks, CA, & London, UK: Sage.
- Newman, J. W. (1995). Socioeconomic class and education: In what ways does class affect the educational process? In J. L. Kincheloe & S. R. Steinberg (Eds.), *Thirteen questions: Reframing education's conversation* (2nd ed., pp. 193-203). New York, NY: Peter Lang.
- Nguyen, N. (2010). *Early post-school outcomes of indigenous youth: The role of literacy and numeracy*. National Centre for Vocational Education Research Briefing Paper 2308. Retrieved from <http://www.ncver.edu.au/publications/2308.html>
- Nieto, S. (2005). Public education in the twentieth century and beyond: High hopes, broken promises, and an uncertain future. *Harvard Educational Review*, 75(1), 43-64.
- Nucci, L. (2006). Classroom management for moral and social development. In C. Evertson & C. Weinstein (Eds.), *Handbook of classroom management: Research, practice, and contemporary issues* (pp. 711-731). New York, NY: Routledge.
- Ochoa, G., & Pineda, D. (2008). Deconstructing power, privilege, and silence in the classroom. *Radical History Review*, 102, 45-62.
- OECD (2000). *Measuring student knowledge and skills: The PISA 2000 assessment of reading, mathematical and scientific literacy*. Paris, France: PISA.
- OECD (2009). *Equally prepared for life? How 15-year-old boys and girls perform in school*. Retrieved from <http://www.oecd.org/dataoecd/59/50/42843625.pdf>
- OECD (2010). *PISA 2009 results: Executive summary*. Retrieved from www.oecd.org
- O'Farrell, C. (1999). Postmodernism for the uninitiated. In D. Meadmore, B. Burnett, & P. O'Brien (Eds.), *Understanding education: Contexts and agendas for the new millennium* (pp. 11-17). Sydney, NSW: Prentice Hall.
- Ognyanova, K. (2010). Talking past each other: Academic and media framing of literacy. *Digital Culture & Education*, 2(1), 44-61.
- Ohanian, S. (1999). *One size fits few: The folly of educational standards*. Portsmouth, NH: Heinemann.
- Olsen, B., & Sexton, S. (2009). Threat rigidity, school reform, and how teachers view their work inside current education policy contexts. *American Educational Research Journal*, 46(1), 9-44.
- O'Mara, J., & Gutierrez, A. (2010). Classroom teachers as co-researchers: The affordances and challenges of collaboration. *Australian Journal of Language and Literacy*, 33(1), 41-53.

- Park, H-Y. (2008). "You are confusing!": Tensions between teacher's and students' discourses in the classroom. *The Journal of Classroom Interaction*, 43(1), 4-13.
- Parker, W. (2006). Public discourse in schools: Purposes, problems, possibilities. *Educational Researcher*, 35(8), 11-18.
- Parr, G., & Bellis, N. (2006). "To be strictly educated"? Learning and teaching in an age of neo-liberal agendas. *English in Australia*, 41(1) 7-17.
- Patty, A. (2010, February 17). Better supervision of tests needed to stop cheats, say schools. *Sydney Morning Herald*. Retrieved from <http://www.smh.com.au/>
- Payne, G., & Payne, J. (2004). *Key concepts in social research*. London, UK, & Thousand Oaks, CA: Sage.
- Pegg, J., & Panizzon, D. (2007). Inequities in student achievement for literacy: Metropolitan versus rural comparisons. *Australian Journal of Language and Literacy*, 30(3), 177-190.
- Pelletier, L., & Sharp, E. (2009). Administrative pressures and teachers' interpersonal behaviour in the classroom. *Theory and Research in Education*, 7(2), 174-183.
- Phelps, R. (2003). *Kill the messenger: The war on standardized testing*. New Brunswick, NJ: Transaction.
- Phelps, R. (Ed.). (2005). *Defending standardized testing*. Mahwah, NJ: Lawrence Erlbaum.
- Phillips, K. (2004). *Testing controversy: A rhetoric of educational reform*. Cresskill, NJ: Hampton Press.
- Phillips, L., & Jørgensen, M. (2002). *Discourse analysis as theory and method*. London, UK: Sage.
- Phillips, N., & Hardy, C. (2002). *Discourse analysis: Investigating processes of social construction*. Thousand Oaks, CA: Sage.
- Pianta, R. C. (2006). Classroom management and relationships between children and teachers: Implications for research and practice. In C. M. Evertson & C. S. Weinstein (Eds.), *Handbook of classroom management: Research, practice, and contemporary issues* (pp. 685-709). New York, NY: Routledge.
- Popham, W. J. (2001). *The truth about testing: An educator's call to action*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Popham, W. J. (2003a). Preparing for the coming avalanche of accountability tests. In Harvard College (Ed.), *Spotlight on high-stakes testing* (pp. 9-15). Cambridge, MA: Harvard Education Press.
- Popham, W. J. (2003b). Are your state's tests instructionally sensitive? In Harvard College (Ed.), *Spotlight on high-stakes testing* (pp. 17-22). Cambridge, MA: Harvard Education Press.

- Popham, W. J. (2007). Instructional insensitivity of tests: Accountability's dire drawback. *The Phi Delta Kappan*, 89(2), 146-150, 155.
- Prashad, V. (2006). The global war against teachers. *Radical History Review*, 95, 9-20.
- Punch, K. (1998). *Introduction to social research: Quantitative and qualitative approaches*. London, UK, & Thousand Oaks, CA: Sage.
- Purpel, D. (1999). *Moral outrage in education*. New York, NY: Peter Lang.
- Ravitch, D. (2003). A historic document. In P. E. Peterson (Ed.), *Our schools and our future ... Are we still at risk?* (pp. 25-37). Stanford, CA: Hoover Institution Press, Stanford University Press.
- Reeve, J. (2006). Extrinsic rewards and inner motivation. In C. M. Evertson & C. S. Weinstein (Eds.), *Handbook of classroom management: Research, practice, and contemporary issues* (pp. 645-664). New York, NY: Routledge.
- Reid, A. (2010). *Accountability and the public purposes of education*. Retrieved from www.aeufederal.org.au/Publications/2010/NS/AReid.pdf
- Reimer, C. (2005). Amplifying student performance by teaching and modeling responsibility. In P. C. Miller (Ed.), *Narratives from the classroom: An introduction to teaching* (pp. 146-163). Thousand Oaks, CA: Sage.
- Rhodes, L. A. (2007). Is there a standard for meeting standards? In Education Week, *The last word: The best commentary and controversy in American education* (pp. 54-59). San Francisco, CA: John Wiley & Sons, Jossey-Bass.
- Riehl, C. (2001). Bridges to the future: The contributions of qualitative research to the sociology of education. *Sociology of Education*, 74 (Extra Issue), 115-134.
- Rishel, T. (2005). Maneuvering the emotional and social demands of the classroom. In P. C. Miller (Ed.), *Narratives from the classroom: An introduction to teaching* (pp. 164-182). Thousand Oaks, CA: Sage.
- Rishel, T., & Miller, P. C. (2005). Positive relationships for effective teaching. In P. C. Miller (Ed.), *Narratives from the classroom: An introduction to teaching* (pp. 127-145). Thousand Oaks, CA: Sage.
- Robertson, S. L. (2010). Teachers matter ... Don't they? Placing teachers and their work in the global knowledge economy. In S. Macrine, P. McLaren, & D. Hill (Eds.), *Revolutionizing pedagogy: Education for social justice within and beyond global neo-liberalism* (pp. 235-255). New York, NY: Palgrave Macmillan.
- Rohlen, T. (2002). Concluding observations: Wider contexts and future issues [Epilogue]. In G. DeCoker (Ed.), *National standards and school reform in Japan and the United States* (pp. 177-205). New York, NY: Teachers College Press, Columbia University.

- Rotherham, A. (2008). The translators: The media and school choice research. *The Phi Delta Kappan*, 89(5), 376-379.
- Rothman, R. (2001). One hundred fifty years of testing. In *The Jossey-Bass reader on school reform* (pp. 419-433). San Francisco, CA: Jossey-Bass.
- Rowan, L. (2001). *Write me in: Inclusive texts in the primary classroom*. Newtown, NSW: Primary English Teaching Association.
- Russell, N. U. (2002). The role of the private sector in determining national standards: How *Juku* undermine Japanese Educational Authority. In G. DeCoker (Ed.), *National standards and school reform in Japan and the United States* (pp. 158-175). New York, NY: Teachers College Press, Columbia University.
- Rutledge, S., Harris, D., & Ingle, W. (2010). How principals “bridge and buffer” the new demands of teacher quality and accountability: A mixed-methods analysis of teacher hiring. *American Journal of Education*, 116(2), 211-242.
- Sacks, P. (2000). *Standardized minds: The high price of America's testing culture and what we can do to change it*. New York, NY: Da Capo Press.
- Sawyer, R. K. (2004). Creative teaching: Collaborative discussion as disciplined improvisation. *Educational Researcher*, 33(2), 12-20.
- Saldaña, J. (2009). *The Coding Manual for Qualitative Researchers*. London, UK: Sage.
- Scheurich, J. J., & McKenzie, K. B. (2005). Foucault's methodologies: Archaeology and genealogy. In N. Denzin & Y. Lincoln (Eds.), *The Sage handbook of qualitative research* (3rd ed., pp. 841-868). Thousand Oaks, CA: Sage.
- Schleppegrell, M., Greer, S., & Taylor, S. (2008). Literacy in history: Language and meaning. *Australian Journal of Language and Literacy*, 31(2), 174-187.
- Shaker, P., & Ruitenberg, C. (2007). “Scientifically-based research”: The art of politics and the distortion of science. *International Journal of Research & Method in Education*, 30(2), 207-219.
- Shanker, A. (2007). By all measures: Coming to terms on world-class standards. In Education Week, *The last word: The best commentary and controversy in American education* (pp. 47-50). San Francisco, CA: John Wiley & Sons, Jossey-Bass.
- Sherington, G., & Campbell, C. (2004). Australian liberalism, the middle class and public education from Henry Parkes to John Howard. *Education Research and Perspectives*, 31(2), 59-77.
- Sherman, W. H., & Grogan, M. (2003). Superintendents' responses to the achievement gap: An ethical critique. *International Journal of Leadership in Education: Theory and Practice*, 6(3), 223-237.
- Sim, A. (2006). An investigation of the literacy demands and support given to a Year 8 class. *Australian Journal of Language and Literacy*, 29(3), 240-251.

- Simpson, S. (2010, October). *Naming and shaming for pre-schools*. New South Wales Teachers Federation. Retrieved from http://www.nswtf.org.au/edu_online/143/detpre.html
- Skelcher, C., Mathur, N., & Smith, M. (2005). The public governance of collaborative spaces: Discourse, design and democracy. *Public Administration* 83(3), 573-596.
- Slavin, R. (2004). Education research can and must address “what works” questions. *Educational Researcher*, 33(1), 27–28.
- Slavin, R. (2008). Evidence-based reform in education: Which evidence counts? *Educational Researcher*, 37(1), 47-55.
- Sloane, F., & Kelly, A. (2003). Issues in high-stakes testing programs. *Theory into Practice*, 42(1), 12-17.
- Smith, T., Edwards-Groves, C., & Kemmis, R. B. (2010). Pedagogy, education and praxis [Editorial]. *Pedagogy, Culture & Society*, 18(1), 1-8.
- Snow, C. E., Porche, M. V., Tabors, P., & Harris, S. (2007). *Is literacy enough?: Pathways to academic success for adolescents*. Baltimore, MD: Paul H. Brookes.
- Snyder, I. (2008). *The literacy wars: Why teaching children to read and write is a battleground in Australia*. Sydney, NSW: Allen & Unwin.
- Snyder, I. (2009). The stories that divide us: Media (mis)representations of literacy education. *English in Australia*, 44(1), 13-23.
- Stanley, W. (1995). Socioeconomic class and education: In what ways does class affect the educational process? In J. L. Kincheloe & S. R. Steinberg (Eds.), *Thirteen questions: Reframing education's conversation* (2nd ed., pp. 205-214). New York, NY: Peter Lang.
- Stern, P. N. (2007). On solid ground: Essential properties for growing grounded theory. In A. Bryant & K. Charmaz (Eds.), *The SAGE handbook of grounded theory* (pp. 114-126). London, UK: Sage.
- Stringfield, S., & Yakimowski-Sreblick, M. (2005). Promise, progress, problems, and paradoxes of three phases of accountability: A longitudinal case study of the Baltimore City Public Schools. *American Educational Research Journal*, 42(1), 43-75.
- Stubbs, M. (1998). *Discourse analysis: The sociolinguistic analysis of natural language*. Oxford, UK: Basil Blackwell.
- Suggett, D. (2010, May 18). NAPLAN and the states: An intriguing result. *Inside Story*. Retrieved from <http://inside.org.au/>

- Tankersley, K. (2007). *Tests that teach: Using standardized tests to improve instruction*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Taras, M. (2007). Machinations of assessment: Metaphors, myths and realities. *Pedagogy, Culture & Society*, 15(1), 55-69.
- Taylor, S. (2001a). Locating and conducting discourse analytic research. In M. Wetherell, S. Taylor, & S. Yates (Eds.), *Discourse as data: A guide for analysis* (pp. 5-48). Sydney, NSW: Sage.
- Taylor, S. (2001b). Evaluating and applying discourse analytic research. In M. Wetherell, S. Taylor, & S. Yates (Eds.), *Discourse as data: A guide for analysis* (pp. 311-330). Sydney, NSW: Sage.
- Tharp, J. M. (2008). *Breaking the cycle of failed school reform: What five failed reforms tell us*. Lanham, MD: Rowman & Littlefield Education.
- Thomson, S., & Buckley, S. (2007). *Informing mathematics pedagogy: TIMSS 2007 Australia and the world*. Retrieved from http://www.acer.edu.au/documents/TIMSS_2007InformingMathsPedagogyreport.pdf
- Thwaite, A., & Rivalland, J. (2009). How can analysis of classroom talk help teachers to reflect on their practices? *Australian Journal of Language and Literacy*, 32(1), 38-54.
- Tomasevski, K. (2006). *The state of the right to education worldwide: Free or fee: 2006 global report*. [Online]. Retrieved from http://www.katarinatomasevski.com/images/Global_Report.pdf
- Torgerson, C. (2006). Publication bias: The Achilles' heel of systematic reviews? *British Journal of Educational Studies*, 54(1), 89-102.
- Torok, S., McMorris, R., & Lin, W-C. (2004). Is humor an appreciated teaching tool? Perceptions of Professors' teaching styles and use of humor. *College Teaching*, 52(1), 14-20.
- Tucker, M. S., & Coddling, J.-B. (2001). Setting high standards for everyone. In *The Jossey-Bass reader on school reform* (pp. 470-482). San Francisco, CA: Jossey-Bass.
- Tyack, D., & Cuban, L. (1995). *Tinkering toward Utopia: A century of public school reform*. Cambridge, MA: Harvard University Press.
- Ullucci, K. (2009). "This has to be family": Humanizing classroom management in urban schools. *The Journal of Classroom Interaction*, 44(1), 13-28.
- Valli, L., & Buese, D. (2007). The changing roles of teachers in an era of high-stakes accountability. *American Educational Research Journal*, 44(3), 519-558.
- Van Blerkom, M. L. (2009). *Measurement and statistics for teachers*. New York, NY: Routledge.

- Van Petegem, P., Creemers, B. P. M., Rossel, Y., & Aelterman, A. (2005). Relationships between teacher characteristics, interpersonal teacher behaviour and teacher wellbeing. *The Journal of Classroom Interaction*, 40(2), 34-43.
- Walker, J. M. T., & Hoover-Dempsey, K. V. (2006). Why research on parental involvement is important to classroom management. In C. M. Evertson & C. S. Weinstein (Eds.), *Handbook of classroom management: Research, practice, and contemporary issues* (pp. 665-684). New York, NY: Routledge.
- Walker, R., & Adelman, C. (1976). Strawberries. In M. Stubbs & S. Delamont (Eds.), *Explorations in classroom observation* (pp. 133-150). New York, NY: John Wiley & Sons.
- Wardekker, W. L. (2000). Criteria for the quality of research. *Mind, Culture and Activity*, 7(4), 259-272.
- Warmington, P., & Murphy, R. (2004). Could do better? Media depictions of UK educational assessment results. *Journal of Educational Policy*, 19(3), 285-299.
- Warne, B. M. (2006). Teaching conventions in a state-mandated testing context. *The English Journal*, 95(5), 22-27.
- Wearmouth, J. (2008). Testing, assessment and literacy learning in schools: A view from England. *Curriculum Perspectives*, 28(3), 77-81.
- Webb, P. T. (2005). The anatomy of accountability. *Journal of Education Policy*, 20(2), 189-208.
- Webb, P. T. (2006). The choreography of accountability. *Journal of Education Policy*, 21(2), 201-214.
- Weedon, C. (1987). *Feminist practice and poststructuralist theory*. Cambridge, MA: Basil Blackwell.
- Weimer, M. (2006). *Enhancing scholarly work on teaching and learning: Professional literature that makes a difference*. San Francisco, CA: Jossey-Bass.
- Wentzel, K. R. (2006). A social motivation perspective for classroom management. In C. M. Evertson & C. S. Weinstein (Eds.), *Handbook of classroom management: Research, practice, and contemporary issues* (pp. 619-643). New York, NY: Routledge.
- Whipp, R. (1998). Qualitative methods: Technique or size? In K. Whitfield & G. Strauss (Eds.), *Researching the world of work: Strategies and methods in studying industrial relations* (pp. 51-63). Cornell, NY: ILR Press.
- Whitehead, D. (2008). Testing like you teach: The challenge of constructing local, ecologically valid tests. *English Teaching: Practice and Critique*, 7(3), 10-25.

- White-Smith, K., & White, M. (2009). High school reform implementation: Principals' perceptions on their leadership role. *Urban Education*, 44(3), 259-279.
- Whitty, G. (2006). Education(al) research and education policy making: Is conflict inevitable? *British Educational Research Journal*, 32(2), 159-176.
- Whitty, G., & Wisby, E. (2007). Whose voice? An exploration of the current policy interest in pupil involvement in school decision-making. *International Studies in Sociology of Education*, 17(3), 303-319.
- Wilde, S. (2002). *Testing and standards: A brief encyclopedia*. Portsmouth, NH: Heinemann.
- William, D. (2008). What should education research do, and how should it do it? *Educational Researcher*, 37(7), 432-438.
- Willet, L., & Gardiner, A. (2009, July 9-12). *Testing Spelling – Exploring NAPLAN*. Lecture presented at the Australian Literacy Educators Association Conference, Hobart, TAS. [Transcript]. Retrieved from <http://www.englishliteracyconference.com.au/files/documents/hobart/conferencePapers/nonRefereed/Willet-Gardiner.pdf>
- Williamson, J., & Myhill, M. (2008). Under “constant bombardment”: Work intensification and the teachers' role. In D. Johnson & R. Maclean (Eds.), *Teaching: Professionalization, development and leadership* (pp. 25-43). Dordrecht, Netherlands: Springer.
- Wilson, B., Croxson, B., & Atkinson, A. (2006). “What gets measured gets done”: Headteachers' responses to the English Secondary School Performance Management System. *Policy Studies*, 27(2), 153-171.
- Witkin, S. (2000). An integrative human rights approach to social research. In C. Truman, D. Mertens, & E. Humphries (Eds.), *Research and inequality* (pp. 205-219). London, UK: University College of London Press.
- Woodward, C. (2000). Hearing voices? Research issues when telling respondents' stories of childhood sexual abuse from a feminist perspective. In C. Truman, D. Mertens, & E. Humphries (Eds.), *Research and inequality* (pp. 37-51). London, UK: University College of London Press.
- Woods, A. (2007). What's wrong with benchmarks? Answering the wrong questions with the wrong answers. *Curriculum Perspectives*, 27(3), 1-10.
- Wright, J. (2003). *Poststructural methodologies – The body, schooling and health*. Retrieved from <http://ro.uow.edu.au/edupapers/55>
- Wyatt-Smith, C. (1998). Standardised testing – In whose interests? *English in Australia*, 122, 89-93.
- Wyatt-Smith, C. (1999). Reading for assessment: How teachers ascribe meaning and value to student writing. *Assessment in education: Principles, policy and practice*, 6(2), 195-223.

- Wyatt-Smith, C. (2000). The English/Literacy interface in senior school: Debates in Queensland. *English in Australia*, 127-128, 71-79.
- Wyatt-Smith C. (2008). Literacy testing and quality [Editorial]. *Curriculum Perspectives*, 28(3), 59-64.
- Wyatt-Smith, C., & Campbell, R. (2002). What counts as quality literacy assessment in middle schooling?: Teacher accounts of classroom-based assessment and large-scale testing. *Curriculum Perspectives*, 22(3), 1-11.
- Wyatt-Smith, C., & Castleton, G. (2004). Factors affecting writing achievement: Mapping teacher beliefs. *English in Education*, 38(1), 37-61.
- Wyatt-Smith, C., Castleton, G., Freebody, P., & Cooksey, R. (2003). The nature of teachers' qualitative judgements: A matter of context and salience. Part one: "In-context" judgement. *Australian Journal of Language and Literacy*, 26(2), 11-32.
- Wyatt-Smith, C., & Klenowski, V. (2008). *Examining how moderation is enacted within an assessment policy reform initiative: You just have to learn how to see*. Lecture presented at the 34th International Association for Educational Assessment Annual Conference, Cambridge, UK.
- Wyatt-Smith, C., & Klenowski, V. (2010). Role and purpose of standards in the context of national curriculum and assessment reform for accountability, improvement and equity in student learning. *Curriculum Perspectives*, 30(3), 37-47.
- Yin, R. (2006). Case study methods. In J. Green, G. Camilli, & P. Elmore (Eds.), *Handbook of complementary methods in education research* (pp. 111-122). Mahwah, NJ: Lawrence Erlbaum.
- Zion, S. (2009). Systems, stakeholders, and students: Including students in school reform. *Improving Schools*, 12(2), 131-143.
- Zyngier, D. (2009). Teachers under the pump – and over a barrel – and up the creek: Reframing the current debates about "quality" of education. *Curriculum Perspectives*, 29(1), 83-90.

Appendix

Appendix A: Approval Documentation

A1: Approvals: University of Tasmania Human Research Ethics Committee and Department of Education (Tasmania).

A1-a: University of Tasmania Human Research Ethics Committee approval letter.

A1-b: Department of Education (Tasmania) approval letter.

A2: Preliminary contact letter to principal.

A3: Information letters provided to principal and teachers.

A3-a: Information letter provided to principal.

A3-b: Information letter provided to teachers.

A3-c: Information letter provided to students.

A4: Consent form provided to principal and teachers.

Appendix B: Data Instruments

B1: Teacher questionnaires.

B1-a: Questionnaire given to TT, T2 and P.

B1-b: Questionnaire given to T1.

B2: Student questionnaires.

B3: Teacher interview programme.

Appendix C: Data Extracts

C1: Extract from researcher's field notes.

C2: Coded extract from interview transcript.

C3: Coded extract from testing session transcript.

C4: Student questionnaire responses.

Appendix D: List of Codes by Frequency

D1: Codes by frequency: Interview transcripts.

D2: Codes by frequency: Testing session transcripts.

Appendix A: Approval Documentation

A1: Approvals: University of Tasmania Human Research Ethics Committee and Department of Education (Tasmania).

A1-a: University of Tasmania Human Research Ethics Committee approval letter.

MEMORANDUM	Private Bag 01 Hobart Tasmania 7001 Australia Telephone (03) 6226 2764 Facsimile (03) 6226 7148 Marilyn.Knott@utas.edu.au http://www.research.utas.edu.au/index.htm	
HUMAN RESEARCH ETHICS COMMITTEE (TASMANIA) NETWORK		

MINIMAL RISK ETHICS APPLICATION APPROVAL

27 February 2006

Dr Claire Hiller
 School of Education
 Private Bag 66
 Hobart

Ethics reference: H8730
'Standardised literacy testing: Issues, approaches, effects'.
PhD candidate: Lauren Johnson

Dear Dr Hiller

Acting on a mandate from the Tasmania Social Sciences HREC, the Chair of the committee considered and approved the above project on 24 February 2006.

All committees operating under the Human Research Ethics Committee (Tasmania) Network are registered and required to comply with the *National Statement on the Ethical Conduct in Research Involving Humans 1999* (NHMRC guidelines).

Therefore, the Chief Investigator's responsibility is to ensure that:

- 1) All researchers listed on the application comply with HREC approved application.
- 2) Modifications to the application do not proceed until approval is obtained in writing from the HREC.
- 3) The confidentiality and anonymity of all research subjects is maintained at all times, except as required by law.
- 4) Clause 2.37 of the National Statement states:
An HREC shall, as a condition of approval of each protocol, require that researchers immediately report anything which might warrant review of ethical approval of the protocol, including:
 - a) *Serious or unexpected adverse effects on participants;*
 - b) *Proposed changes in the application; and*
 - c) *Unforeseen events that might affect continued ethical acceptability of the project.*


The report must be lodged within 24 hours of the event to the Ethics Executive Officer who will report to the Chairs.

A PARTNERSHIP PROGRAM IN CONJUNCTION WITH THE DEPARTMENT OF HEALTH AND HUMAN SERVICES

A1-1 (continued)

- 5) All participants must be provided with the current Information Sheet and Consent form as approved by the Ethics Committee.
- 6) The Committee is notified if any investigators are added to, or cease involvement with, the project.
- 7) This study has approval for four years contingent upon annual review. An *Annual Report* is to be provided on the anniversary date of your approval. Your first report is due [12 months from 'Ethics Committee Approval' date]. You will be sent a courtesy reminder by email closer to this due date.
Clause 2.35 of the National Statement states:
As a minimum an HREC must require at regular periods, at least annually, reports from principal researchers on matters including:
 - a) Progress to date or outcome in case of completed research;
 - b) Maintenance and security of records;
 - c) Compliance with the approved protocol, and
 - d) Compliance with any conditions of approval.
- 8) A *Final Report* and a copy of the published material, either in full or abstract, must be provided at the end of project.

Yours sincerely



Ethics Executive Officer

A1-b: Department of Education (Tasmania) approval letter.

Department of Education
EDUCATIONAL PERFORMANCE SERVICES
2/99 Bathurst Street, Hobart
GPO Box 169, Hobart, TAS 7001 Australia



File: 650295

20 July 2007

Dr Claire Hiller
Senior Lecturer
School of Education
University of Tasmania
Private Bag 66
HOBART TAS 7000

Dear Dr Hiller

Standardised Literacy Testing: Issues, Approaches, Effects – Ms Lauren Johnson

I have been advised by the Educational Performance Research Committee that the above research adheres to the guidelines established and therefore have no objection to approving this study retrospectively. In order to update our records it would be appreciated if you could forward to us a brief outline of your progress to date.

We note that you sought permission from the Principal of [REDACTED] High School prior to your interviews with the English/Literacy teachers and that you conducted all interviews in an appropriate manner.

A copy of your final report should be forwarded to Patricia Lloyd, Educational Performance Services, Department of Education, GPO Box 169, Hobart 7001 at your earliest convenience and within six months of the completion of the research phase in Department of Education schools.

Yours sincerely

Manager
(Educational Performance Services)
Cc Ms Lauren Johnson

A2: Preliminary contact letter to principal.

[Faculty letterhead]

18 July, 2006

[name]

Principal

[Address, line 1]

[Address, line 2]

[Address, line 3]

Dear [P],

I am writing to ask for your permission to undertake the following research in your school for the requirements of obtaining my PhD. The research entitled *Standardised Literacy Testing: Issues, Approaches, Effects* is being undertaken as part of my research study to meet requirements for my PhD. I am hoping to study only one secondary school in my research. Of all those English/Literacy teachers responsible for standardised literacy testing at this school, I intend to interview, observe and study as many as are willing.

I am seeking your permission for research to be undertaken with willing English/Literacy teachers in your school. The purpose of the research is to gain a critical understanding of teachers' perspectives on and practices with standardised literacy testing in the classroom; and to investigate how these might affect the ways the tests are approached by their students and, in turn, the responses and data collected in the standardised literacy tests. Background information about the topic and the research process will be provided to the participants prior to the commencement and signed consent will be sought. Participation will involve: teachers will complete

anonymous questionnaires, engage in individual interviews, and agree to observation of classroom teaching and testing; student participants will complete anonymous questionnaires and also be the subjects of observation during classroom testing. Interviews will be audio taped then transcribed; observational notes will be taken from classroom observation.

Every precaution will be taken to minimise any risks that may cause the school and the participants involved the possibility of emotional distress, anxiety or embarrassment. Confidentiality of records will be maintained and established research protocols respected. Where data is disclosed in the thesis, name and other identifiers for the school and participants will not be used. This research will only focus on the perspectives on and practices with standardised literacy tests in the classroom, and not teaching practice or information of a personal nature for either teachers or students. You may withdraw your permission to allow the research to be conducted at any time and you will not be prejudiced.

With respect to the time commitment required of participant teachers and students, the student questionnaire will take no more than ten minutes to complete, the teacher questionnaire no more than ten minutes, and teacher interviews approximately one hour. Time frames depend on each participant but extra time is not expected nor necessary.

Should you agree to allow the research study in your school, your assistance is sought for the following specific activities. As Principal, you could support the research study through the role of disseminating information as first stage in the recruitment of participating teachers and students. I would welcome your help with this promotion and invitation stage of the research and you may have suggestions to offer with regards the distribution of information to the school community and potential participants. I have attached background information sheets for teachers and students. Also with your permission, I have attached an introductory

article written for the benefit of the parent body and suitable for inclusion in the school newsletter. Parents and guardians who do **not** wish their child/ren to be involved could notify the school of their decision.

More details concerning the research topic and research methodology are contained in the attached information sheets. I would be pleased to provide any further information you require to help your assessment of this request; alternately you can contact directly my supervisor, Dr. Claire Hiller, Senior Lecturer, Faculty of Education, University of Tasmania. Her contact number is (03) 6226 2560. Additionally, more detail regarding the study is found in the information sheets and consent forms attached to this letter. The research has received ethical approval from the Human Research Ethics Committee (Tas) Network and the Department of Education, Tasmania. If you have concerns of an ethical nature or complaints about the manner in which the project is conducted, you may contact the Executive Officer of the Network, Amanda McAully, (03) 6226 2763.

Please feel free to contact me without delay with your response to this request. Otherwise, I will telephone you within a fortnight of sending this letter to discuss the research and conclude whether or not your school will be participating in this project.

Yours sincerely,

Ms. Lauren Johnson, PhD Candidate.

A3: Information letters provided to principal and teachers.

A3-a: Information letter provided to principal.

Background Information Sheet

Standardised Literacy Tests

Australia has defined benchmarks of literacy and numeracy standards for students in primary and secondary schools. These benchmarks represent the minimum acceptable standard without which a student will have difficulty making sufficient progress at school. The Ministerial Council on Education, Employment, Training and Youth Affairs (MCEETYA) has as its goal that all students will achieve at least the benchmark level of performance; as a result, students are tested during years three, five, seven and nine, and the results are gathered and reported. These tests are known by several names, for instance, 'standardised', 'benchmarking' and 'norm-referenced' tests. A standard, benchmark or norm is devised and all responses and tests are compared with this. Publication of benchmark results enables the monitoring of student progress towards the attainment of the goals set by MCEETYA.

Research Project

Year seven and nine students at your school are invited to participate in a research project undertaken by Ms. Lauren Johnson, as part of the requirements to obtain a PhD. The proposed research is an examination of English/Literacy teachers' perceptions and practices of standardised literacy testing and how these might affect the ways students and schools approach such tests. The school has already given permission for the research investigation to take place.

As part of the study, Ms Johnson plans to observe and take notes on what actually happens in the classroom when standardised literacy testing is taking place – teachers' and students' reactions to the tests, language used, actions and behaviours, for instance – and invites observed students to fill in an anonymous questionnaire about their experiences of standardised literacy testing. The questionnaires do not ask for names, and in being observed students' identities will not be disclosed in the thesis. The questionnaire will take you no more than five minutes to complete. The research

project will not gather data about outcomes or other longer term uses made of the findings or results of standardised literacy tests. Nor will the researcher be gathering personal information about students' backgrounds or level of educational progress.

Please inform the school principal should you wish for your child/ren to **not** be involved in this study.

To discuss this research project and any concerns relating to the research methods, or for further information or queries, please contact the researcher direct:

Ms. Lauren Johnson, Investigator
School of Education (Hobart)
Private Bag 66
Hobart TAS 7001
Phone: (03) 6226 7239
Email: Lauren.Johnson@utas.edu.au
<http://fcms.its.utas.edu.au/educ/educ/index.asp>

The following websites contain further information about standardised literacy tests:

Ministerial Council on Education, Employment, Training and Youth Affairs:
<http://www.mceetya.edu.au/>

Federal Department of Education, Science and Training: <http://www.dest.gov.au/>

Department of Education, Tasmania: <http://www2.education.tas.gov.au/>

Curriculum Corporation: <http://www.curriculum.edu.au/>

A3-b: Information letter provided to teachers.

Information Sheet for Teachers

Title of Investigation: *Standardised Literacy Testing: Issues, Approaches, Effects*

Name of Chief Investigator: Dr. Claire Hiller, Senior Lecturer, Faculty of Education,
University of Tasmania, Hobart.

Name of Investigator: Ms. Lauren Johnson.

Date: 18 July 2006

You are invited to participate in a research project being undertaken by Ms. Lauren Johnson, as part of the requirements to obtain a Doctorate of Philosophy, in the Faculty of Education. The proposed research is an examination of English/Literacy teachers' perceptions and practices of standardised literacy testing and how these might affect the ways students and schools approach such tests.

Inclusion and exclusion criteria

The Principal of your school has agreed for your school to be involved and all teachers involved in the school's year seven and nine English/Literacy teaching programme are invited to participate as volunteers in the research project.

Participant teachers will be recruited via the school Principal. The Principal will distribute information sheets and consent forms to teachers of year seven and nine English/Literacy who will or might be required to use standardised literacy tests during the 2006 school year.

Purpose of the study; Participant benefit

Your participation in the research will contribute to education practices by allowing your experiences during standardised literacy testing instances to highlight the issues that are most relevant to teachers in the area of English/Literacy. It is hoped that the study will contribute to your own

personal development as a teacher by providing you with an opportunity to examine and articulate your own assessment pedagogy and practices.

As part of the study, the researcher will conduct anonymous questionnaires of and interviews with English/Literacy teachers, and observe and take notes on the classroom reality of English/Literacy teaching (for the researcher's understanding of the classroom context and embedded English/Literacy assessment practices) and standardised literacy testing. Anonymous questionnaires of students will also be conducted, about their experiences of standardised literacy testing. The questionnaires and observation of students in this research will in no way identify or make identifiable any students.

Timetables and process; Study procedures

As part of the study, the researcher will conduct questionnaires and interviews with English/Literacy teachers to examine teachers' perspectives of and practices with standardised literacy testing. The order of research activities will be as follows for participants.

Once the participant teachers have been established, the teacher questionnaire, lesson and testing observations, student questionnaires, and teacher interviews will take place, respectively.

Lesson observation is entirely voluntary, and the researcher's role in the lessons and frequency of observation will be decided wholly by willing teacher/s. These lessons will be audio-taped and transcribed (as per the teacher interviews), will be anonymous and will only be used by the researcher to contextualise standardised literacy testing observations and in order to provide greater depth in the interview stage.

Teacher questionnaires will be provided by the Principal, and collected by the researcher. The questionnaire consists of eight questions, and is expected to take no more than ten minutes to complete. The questionnaire seeks basic

information about the respondent, such as length of English/Literacy teaching career and exposure to standardised literacy testing, to more detailed information about individual perspectives on standardised literacy testing and the ways the tests are used in the classroom. Questionnaire responses will provide initial data to contextualise that data which will be collected from standardised literacy testing observation and teacher interviews.

Following the questionnaire stage of the research, during the year seven and nine standardised literacy testing classroom sessions at the school, the researcher will observe testing sessions of approximately 90 minutes. The researcher will observe as many of these testing sessions as possible, depending on the testing timetable and scheduling, to be ascertained closer to the time. Immediately after or in a later lesson, student questionnaires will be distributed and collected by the researcher, which are expected to take no more than five minutes to complete. These ask for information in students' experiences of standardised literacy testing and how they believe these tests are used in the classroom.

Teacher interviews will be the final stage of school data collection. Interviews are not expected to take more than an hour each, and the time and place of these will be negotiated between the researcher and individual participants after the observation stage. The research project is allowing for a total of four weeks for the completion of all interviews at this stage. Interviews will be audio-taped, then transcribed and analysed. The topic of interviews will be more in-depth than that of the teacher questionnaires; the list of proposed questions is attached. A transcript will be returned to participants within four weeks for approval following the interview. This process of finalising agreed interview records will be completed within two weeks.

Protocols and protections

Responses and information will be available only to the participant in question. All efforts will be made to maintain your anonymity. You will not be

individually identified in material cited in research reports and publications; teacher and student names will either be substituted or omitted, no other potentially identifiable information – such as specifics of appearance – will be included, and the identity of the school will not be disclosed. You will be given transcripts of your interviews and classroom standardised literacy testing observations for you to edit or change if you wish to do so.

The research data will be stored in a locked cabinet in a locked room at the Faculty of Education, University of Tasmania, Hobart, for a period of at least five years. The data will be destroyed after this period by shredding.

Every precaution will be taken during this research to minimise any risks that may cause participants the possibility of emotional distress, anxiety or embarrassment by providing you with all the necessary details that are relevant to you in this study. The researcher will respect your position as a teacher and will behave towards you and others you are in contact with in a professional, courteous manner. Your participation in the research will not be professionally prejudicial to you. Your responses will not be provided to any other parties, and will not be discussed with any teaching staff or students.

Voluntariness

Your participation in the study is entirely voluntary, and evidenced by signing a consent form. In any case, you are entitled to withdraw from the study at any time without prejudice, and withdraw any data you have supplied to date. Further information and questions can be directed to my supervisor and Chief Investigator for the research project, Dr. Claire Hiller, Senior Lecturer, Faculty of Education, University of Tasmania, Hobart. Her contact number is (03) 6226 2560. The research has received ethics approval from the Human Research Ethics Committee (Tasmania) Network and the Department of Education, Tasmania. If you have any concerns of an ethical nature or complaints about the manner in which the research project is conducted, you

may contact the Executive Officer of the Network, Amanda McAully, (03) 6226 2763.

Your participation in the research study will allow you to be given the opportunity to be informed of the overall results of the study at its conclusion, or of any significant findings during the course of the study. You will also be given copies of the information sheet and a statement of informed consent to keep.

Should you desire involvement in this research project, please read and sign the consent form provided and place in an individual envelope (provided) to be collected by the researcher from the school. There is no need to provide any identifiable information on the envelope itself. I sincerely welcome your indication of interest in this research.

Dr. Claire Hiller
Chief Investigator

Ms. Lauren Johnson
Investigator

Teacher Interview Questions

Do you follow any procedures etc. when 'testing' students, such as repeating instructions or questions, providing help or advice to students, or following only the provided guidelines? Why / why not?

For what reason do you do this? (Who, why, in what ways, possible consequences?)

Does using these mandated literacy tests put any kind of pressure on you to get through the set syllabus/curriculum/topics/materials?

If so, what effects does this have on the ways you introduce the tests into your classroom?

Are you able to identify areas where students might need further help in their learning progress from these tests?

- If so, how do you identify and follow-up on problem areas?
- If not, do you think you should be able to do so through these tests?

Do you think these tests do/should serve a different purpose?

Do you behave in a way while introducing or concluding etc. these tests that you feel supports or undermines the tests':

- Purpose
- Educational merit
- Assessment merit

If yes, for what reason/s do you do this?

Do you think this is widespread in the teaching community?

A3-c: Information letter provided to students.

Information Sheet for Students

Title of Investigation: *Standardised Literacy Testing: Issues, Approaches, Effects*

Name of Chief Investigator: Dr. Claire Hiller, Senior Lecturer, Faculty of Education,
University of Tasmania, Hobart.

Name of Investigator: Ms. Lauren Johnson.

Date: 18 July 2006

You are invited to participate in a research project undertaken by Ms. Lauren Johnson, as part of the requirements to obtain a Doctorate of Philosophy, in the Faculty of Education. The proposed research is an examination of English/Literacy teachers' perceptions and practices of standardised literacy testing and how these might affect the ways students and schools approach such tests.

Purpose of the study

Your participation in the research will contribute to education practices by allowing your experiences during standardised literacy testing instances to highlight the issues that are most relevant to teachers in the area of English/Literacy.

Study procedures

As part of the study, Ms Johnson will observe and take notes on what actually happens in the classroom when English/Literacy assessment and standardised literacy testing are taking place – teachers' and students' reactions to the tests, language used, actions and behaviours, for instance – and invites you to fill in an anonymous questionnaire about your experiences of standardised literacy testing. The questionnaires do not ask for your name, and in being observed your identity will not be disclosed in the thesis. The questionnaire will take you no more than five minutes to complete, and your principal and teacher have allowed for you to participate in the questionnaire.

Your participation in the research will not be prejudicial to you. Your responses will not be provided to any other parties, and it will not be discussed with any teaching staff, students or your parents. The research data will be stored in a locked cabinet in

a locked room at the Faculty of Education, University of Tasmania, Hobart, for a period of at least five years, and then destroyed by shredding.

Voluntariness

Your participation in the study is entirely voluntary, and is not part of your study. You can withdraw from the study at any time without being in any way disadvantaged. For further information and questions on the research, these can be directed to my supervisor, Dr. Claire Hiller, Senior Lecturer, Faculty of Education, University of Tasmania, Hobart. Her contact number is (03) 6226 2560.

Ethics approval

The research has received ethics approval from the Human Research Ethics Committee (Tas) Network and the Department of Education, Tasmania. If you have any concerns of an ethical nature or complaints about the manner in which the research project is conducted, you may contact the Executive Officer of the Network, Amanda McAully, (03) 6226 2763.

I hope you will be willing to help in filling out this questionnaire.

Dr. Claire Hiller
Chief Investigator

Ms. Lauren Johnson
Investigator

A4: Consent form provided to principal and teachers.

Note: 'Teacher' consent form applied to TT, T1, T2 and P. The term 'teacher' on the title differentiated between teacher/principal and student.

CONSENT FORM (teachers)

Standardised Literacy Testing: Issues, Approaches, Effects.

1. I have read and understood the 'Information Sheet' for this study.
2. The nature and possible effects of the study have been explained to me.
3. I understand that the study involves a semi-structured audio-taped & transcribed interview of approximately one hour on the topic of their perspectives and perceived practices of standardised literacy testing, and observation and note-taking of English/Literacy lessons – if I so choose – and standardised literacy testing lessons – approximately two sessions of 90 minutes' duration.
4. I understand that all research data will be securely stored on University of Tasmania premises for a period of five years, and then destroyed.
5. Any questions that I have asked have been answered to my satisfaction.
6. I understand that the researchers will not disclose my identity, and that the data will be used only for the purposes of research.
7. I agree that research data gathered for the study may be published provided that I cannot be identified as a subject.
8. I agree to participate in this investigation and understand that I may withdraw at any time without any effect, and withdraw any data I have supplied to date.

Name of Participant:

Signature:

Date:

Statement by Investigator

☐

I have explained this project and the implications of participation in it to this volunteer and I believe that the consent is informed and that he/she understands the implications of participation

If the Investigator has not had an opportunity to talk to participants prior to them participating, the following must be ticked.

☐

The participant has received the Information Sheet in which my details have been provided so that participants have had opportunity to contact me prior to them consenting to participate in this project.

Name of Investigator

Signature of
Investigator

Appendix B: Data Instruments

B1: Teacher questionnaires.

B1-a: Questionnaire given to TT, T2 and P.

Teacher Questionnaire: Teacher Experienced with Standardised Literacy Tests

1. For how many years have you:
 - a) Been teaching?
 - b) And been teaching English/Literacy/Humanities?
 - c) And used standardised/benchmarking literacy tests in your classroom(s)?
.....

2. Have these tests positively or negatively affected your (please circle):

a) Teaching pedagogy and/or practice	Yes / No / Unsure
b) Classroom / behaviour management	Y / N / U
c) Curriculum / syllabus delivery	Y / N / U
d) Assessment and/or monitoring of student learning progress	Y / N / U
e) Teacher-student, student-student and/or teacher-teacher relations	Y / N / U
f) Other (please specify):	Y / N / U

3. In what ways have they affected your:
 - a) Teaching pedagogy and/or practice
.....
.....
.....
.....
 - b) Classroom / behaviour management
.....
.....
.....
.....
 - c) Curriculum / syllabus delivery
.....
.....
.....
.....
 - d) Assessment and/or monitoring of student learning progress
.....
.....
.....
.....
 - e) Teacher-student, student-student and/or teacher-teacher relations
.....
.....
.....
.....

- f) Other.....
.....
.....
.....

4. Since first implementing mandated literacy tests in your classroom(s), have you developed any professional techniques of usage, introduction, follow-up, etc.? Please detail.

.....
.....
.....
.....
.....
.....

5. In your professional opinion, do you feel such mandated literacy tests could or should be utilised differently (such as in the classroom, or otherwise such as development, reporting, or presentation)? If yes, do you think this affects the ways you use the tests in the classroom? Why / why not?

.....
.....
.....
.....
.....
.....

6. In what ways do you believe such mandated standardised/benchmarking literacy tests either benefit or encumber/make suffer:

- a) Students
.....
.....
.....
- b) Parents
.....
.....
.....
- c) Schools
.....
.....
.....
- d) Teachers
.....
.....
.....
- e) Other

.....
.....
.....

7. What do you see as the place of testing in modern society with trends in education?

.....
.....
.....
.....
.....
.....

8. Any further comments with respect to standardised/benchmarking literacy testing and the ways in which they are used in the classroom.

.....
.....
.....
.....
.....
.....
.....

B1-b: Questionnaire given to T1.

Teacher Questionnaire: Teacher Inexperienced with Standardised Literacy Tests

1. For how many years have you:
 - a) Been teaching?
 - b) And been teaching English/Literacy/Humanities?
 - c) And used standardised/benchmarking literacy tests in your classroom(s)?
.....

2. Do you foresee these tests to positively or negatively affect your (please circle):

a) Teaching pedagogy and/or practice	Yes / No /
Unsure	
b) Classroom / behaviour management	Y /
N / U	
c) Curriculum / syllabus delivery	Y /
N / U	
d) Assessment and/or monitoring of student learning progress	Y /
N / U	
e) Teacher-student, student-student and/or teacher-teacher relations	Y /
N / U	
f) Other (please specify):	Y /
N / U	

3. In what ways do you foresee these tests to affect your:
 - a) Teaching pedagogy and/or practice

 - b) Classroom / behaviour management

 - c) Curriculum / syllabus delivery

 - d) Assessment and/or monitoring of student learning progress

 - e) Teacher-student, student-student and/or teacher-teacher relations

 - f) Other

.....
.....
.....

4. Some teachers develop professional techniques of usage, introduction, follow-up, etc. when implementing mandated literacy tests in their classrooms. Do you believe this might happen to you? If so, do you think this would be deliberate, and what purposes would be served by this? Please detail.

.....
.....
.....
.....
.....
.....
.....

5. In your professional opinion, do you feel such mandated literacy tests could or should be utilised differently (such as in the classroom, or otherwise such as development, reporting, or presentation)? If yes, do you think this affects the ways you could, or will, use the tests in the classroom? Why / why not?

.....
.....
.....
.....
.....
.....
.....

6. In what ways do you believe such mandated standardised/benchmarking literacy tests either benefit or encumber/make suffer:

a) Students

.....
.....
.....

b) Parents

.....
.....
.....

c) Schools

.....
.....
.....

d) Teachers

.....
.....
.....

e) Other

.....

.....
.....

7. What do you see as the place of testing in modern society with trends in education?

.....
.....
.....
.....
.....
.....
.....

8. Any further comments with respect to standardised/benchmarking literacy testing and the ways in which they are used in the classroom.

.....
.....
.....
.....
.....
.....
.....

B2: Student questionnaires.

Student Questionnaires

1. How many of these standardised/benchmarking literacy tests have you done?

.....

2. a) Do you think these tests help you to learn better or allow yourself or your teachers to identify areas of your learning progress that need more attention? Please provide detail for your answer.

.....
.....
.....
.....
.....

- b) Do you or your teachers ever follow up on these?
Why / why not?

.....
.....
.....
.....

If yes, how?

.....
.....
.....
.....

3. Do you take these tests seriously? Please provide detail for your answer.

.....
.....
.....
.....

4. Do you think your teachers take these tests seriously? Please provide detail for your answer.

.....
.....
.....
.....

5. Any additional comments you would like to make about standardised/benchmarking literacy tests or testing, and/or the ways they are used in your classroom.

.....
.....

.....
.....
.....
.....
.....

B3: Teacher interview programme.

Teacher Interview Questions

Do you follow any procedures etc. when 'testing' students, such as repeating instructions or questions, providing help or advice to students, or following only the provided guidelines? Why / why not?

For what reason do you do this? (Who, why, in what ways, possible consequences?)

Does using these mandated literacy tests put any kind of pressure on you to get through the set syllabus/curriculum/topics/materials?

If so, what effects does this have on the ways you introduce the tests into your classroom?

Are you able to identify areas where students might need further help in their learning progress from these tests?

- If so, how do you identify and follow-up on problem areas?
- If not, do you think you should be able to do so through these tests?

Do you think these tests do/should serve a different purpose?

Do you behave in a way while introducing or concluding etc. these tests that you feel supports or undermines the tests':

- Purpose
- Educational merit
- Assessment merit

If yes, for what reason/s do you do this?

Do you think this is widespread in the teaching community?

Appendix C: Data Extracts

C1: Extract from researcher's field notes.

P, Initial Pre-Interview

Thursday, 6 July 2006, 3 pm – 4 pm

- Data from the tests (OER) at [school] formed the ELs norms for the literacy and numeracy standards and progressions.
- The results come out so late in the year that they can only be used mainly for programme planning than for working with students individually about problem areas.
- The results report on reading, writing, overall literacy, and overall numeracy. They used to give more detail, breaking down into strands (those in the Statement and Profiles and other English curricula and syllabi), which was more helpful.
- Indigenous Education Unit: looks at indigenous students performing/scoring below benchmarks.
- The ENI is the numerical value scores for schools: schools given numbers between 0 and 120, usually between 18 and 90. The higher the number, the lower the SES, demographic, and so on.
 - [...] Primary schools ([school's] feeder primary schools) usually get the 80s and 90s.
 - [school and similar schools] high schools usually get ENI of 66.
 - [nearby school] HS scores in the low 50s.
- School size – all comparisons (ask OER)
- 2005 Yr. 7 Monitoring Test Results:

C2: Coded extract from interview transcript.

C2-a: TT, Interview 1, p. 1

TT, Interview 1: 18'51"

Wednesday, 19 July 2006, 9.25-9.45 am

Yr. 9 English/Humanities

TT: Er, give me five minutes to have this conversation, and then I'll have a look at it.

R: Okay, do you mind just holding that?

TT: Holding it? Alright, no trouble. Um – are we recording? Yes. Uh, the short answer is that, every teacher does it differently. So that what I do might be completely different to what the person in the room next door is doing.

R: Yep.

TT: And that, to a large extent, literacy and numeracy tests are the sorts of things that appear in the middle of the year. I do no explicit preparation for literacy tests, um, we use them as a diagnostic tool once the tests have been completed, in that we might look and see that individual X has a particular weakness that we may not have noticed. Um, the other thing we often use them for is that the results are used to confirm or to make us question the ratings we are currently giving [the] students. Other teachers may use them for much deeper diagnostic activities, but in my case I tend to use – they tend to be treated as a bit of an aside, as the sort of thing that you go through the process, and make sure the tests are administered as best you can, and then, if there's something that really jumps out at you, when the results come back, then you will, uh, take some action, according to that information. But, in general terms, they are the sorts of things that we use more as a school-wide indicator of how we're progressing, rather than as an individual diagnostic activity. I think that pretty much sums it up!

R: [chuckles]

TT: The other thing you need to know is that we're dealing with a client-base here where work ethic isn't high, in a lot of the kids. What you're seeing at the moment is really good ... but you're also seeing some individuals with really, really uh ... a heterogeneous mix of individuals, uh, from the very brightest to those who are still at early primary school level – with their literacy – all in one room. Which means that it's particularly difficult to stand up the front of the

C2-b: T1, Interview 1, p. 1

T1, Interview 1

Wednesday, 15 November 2006

- T1: Yes, it seemed like that's the first time they'd had to do one of these, they were used to calling out in class 'what's the answer? What's the answer?' They want you to tell them the answer, and when you tell them you couldn't and they – it was quite distressing for some of them. And it was really quite difficult. Other kids, a few in the grade 7 class which I did – I think the grade 7s handled it better than the grade 9s, in fact. We had a couple of anxiety attacks from a very capable student, who – in the writing exam, one young man, who is a very capable student, very capable – almost to the stage where he couldn't write because of the pressure to 'come up with something, *now*'. But he wrote a page and a half in the end, and I read it afterwards, it was very good. I saw the results, the test results came through the other day, he got an appropriate standard for that, I think it was a 5 Lower or something: he did well. But the anxiousness, the anxiety, brought on by the fact that 'I have to perform, well, and to a timeframe'. They just found that really hard.
- R: Yeah. In the transcripts, one of the testing sessions that I was in, I can't remember which grade it was, but they were really struggling with the concept that they had to do it in this certain amount of time. That they had this enforced-creativity, hour-and-a-half sort of thing, and they were saying, 'what happens if I don't finish it?' And they were told, 'well, you have to finish it'. 'But what if I can't, what if I can't?!' And they did manage to finish it, but I don't think they were very impressed by having those sort of limits pushed on them, that they had to do it, in that time.
- T1: I think one of the things that got me about it, I'll just refer back to something I did in class. We did a writing exercise, a spontaneous exercise, I'd put four words on the board, a noun, an adjective, and perhaps another noun and another adjective, and they had to write for twenty minutes, using those four words. And the first question that I would get, and I've done that in each of my classes ... 'how much do you want?' 'What do you want, and then I'll try and supply it.' ... trying to

C2-c: T2, Interview 1, p. 1

T2, Interview 1

Wednesday, 15 November 2006

[brief unrelated conversation between T2 and R prior to interview conversation]

R: I know that the questionnaire's a bit long and very detailed, but anything that you're ... that's fine.

T2: *[the audio-recorder is small]*

R: I know. I've got a smaller one at home, but it's so small that it just kind of gets lost in my bag. I've got a little list of interview questions, and some of them overlap from the questionnaire. The first one is just asking about procedures that you might follow when you're doing the tests in the classroom. I know that some teachers follow the instructions perfectly, thinking that that's well what they're supposed to do or some teachers might go over some things a bit more.

T2: Yeah, it's actually – I mean, they give you quite a lot of things, especially for the writing questions, that you go through to actually prepare them, the problem being of course that when you have three different tests to do and the numeracy tests and they're all supposed to be done within the one week, you could actually spend a whole lesson on that preparation stuff but you – the reality is that you can't really do that. There are things that I do which I'm sure aren't part of the

procedure, like, uh, sometimes I would read the word out to a kid who isn't sure if it's the word they think it is – I don't think you're really supposed to do that. But actually I think I only did that with the PAT – with the vocabulary tests, so that's a different set of tests than [you've/we've] just been doing. Some – there was something about – if you followed the instructions ... reading/comprehension – one of them was really not organised, I think it was that one, that if you just did what it says, the kids would have been quite confused, I think. So I don't follow every single thing to the letter.

R: But when you're not following it to the letter, you're not doing that because um – you'd be more likely to do that because you know what the students need to be able to –

1, 11, 44

1, 11, 14, 22,

26, 42, 43

C2-d: P, Interview 1, p. 1

P, Interview 1

Wednesday, 15 November 2006

- P: You came at an interesting time. The results have just been released last week and I'm in the process of sorting them out.
- R: That's very good timing for me!
- P: Yeah, it is. So what we're doing is – although the only thing that I've distributed to staff so far is this document, I don't know if [TT]'s shown it to you or not. 24
- R: Oh yeah, he showed me that.
- P: Which is an interesting piece of documentation in that by some strange process they've somehow correlated the performance in the reading, writing and spelling tests, and in the single numeracy test, into some projection towards a performance standard and award, which is a bit of a smoke and mirrors test to me. 24
- R: Well, I asked [TT] how he thought that they –
- P: But funnily enough, some of them are pretty good! 31
- R: Yeah, well! He said that they're pretty close.
- P: Some are, yeah. And the ... at the end which is saying that – that ... there, about the progressions statement, and progressions, and standards, is rather interesting. So it just says that they may not have shown, if you like, competence towards it. And it's significant, I've just picked up a grade nine class at the start of this term and we're doing some extension work, and one of the students, whose first pieces of work I graded at 5 Middle, he's only been assessed at a 4 Upper, so in terms of those kids at the upper end of the scale, it's not bad, but it depends on what sort of day they're having the day of the test, because there are other results I've gotten which are below and above, so it's – OER, to their credit – oh, well, yeah, to their credit – do make the point this is one test. They do make the point. For example, these are the score and percentile tabs. So they're giving a whole stack of this sort of stuff, and all too frequently, I think. Do you want to have a look, do you want these? 3, 4, 7, 9,
20, 25
24, 28,
31
- R: Yeah, that'd be good.

C3: Coded extract from testing session transcript.

C3-a: Year Nine, Writing/Composition, p. 1

Year Nine Writing/Composition Test (session 1 of 2)

Tuesday, 25 July 2006, 9.00-9.45am

52'07" recorded

- TT: ... You've been introduced? You're not sure.
- S: Yeah.
- TT: [9MS1], ... so, it's one of those special things ... okay, you need to ... now, 29
okay!
- S: He's just going out for a smoke!
- TT: That means I don't want to hear another sound – as of now you are under 21
exam conditions, which means that you do not speak – 29
- S: ...
- S: ...
- TT: You do not speak, you do not spend any time worrying about what other 21, 29
people are doing, you are worried only about what you are doing. Now, I am
aware that for some of you, this is going to be a task that you don't enjoy, but 2
everybody in Tasmania has to do this, you're not the only people. You have to 14, 19, 26
do it to the best of your ability. This is not a test that has any influence on the 10
mark you get at the end of the year, it is a test that helps us as a school identify 15, 17
where people are with their literacy – with their reading, writing, spelling –
and helps us with some ideas about how we can help you.
- S: I need another pen. 9
- TT: So it's a really important test from the point of view of helping you. Now I'm
going to give you some instructions, you need to listen really carefully to these 16, 21, 29
instructions, and from now on, there should not be one word. The other really
important thing is to remember that this test is one that will take you possibly
60 or 65 minutes. However, it's quite possible you'll finish before that. You 29, 34
listening [9MS2]?
- S: Yeah.
- TT: If you finish before that, as I said to you about ten minutes ago in tutor, your 16, 29
job is to make sure that you have something sensible that you can be doing.
For most of you, the time is now nine o'clock, we're finishing here at 9.35, 14, 19
and we're gonna come back to it period four, for most of you it's going to be a
situation where you will not get finished during this lesson.

C3-b: Year Seven, Reading, p. 1

Year Seven Reading Test

Tuesday, 8 August 2006, 10.55-11.50am

Yr. 7 Literacy Test 2006: Part A, Reading/Comprehension

T3 and T4: Other teachers

TT: ... who are functionally illiterate. Do you want me to say all that again [*for the tape*]? 2, 14, 19,

R: Okay there you go. 21, 23, 26,

TT: Right. So we've got three kids who are functionally illiterate, and they're gonna 35
take a minute before they realise they can't do it, and then we're going to have to
keep them quiet whilst the others get on with it.

R: Do you want me to take anything? 2, 5, 6, 7,

TT: No, I think I'm right. So, that's a problem, and the other problem is time. There 10
isn't quite enough of it. But we'll do our best. [7FS1], that needs to go for this 29
lesson, please. If I see it, I'll have it. Right you people need to spread yourselves
out, don't want to see people sitting beside each other. 16, 21, 29

[noise and talking]

TT: [R], on my desk is a box of rubbers, would you mind getting about another four?
Please? They're just sitting to the left of the computer.

S: ...

TT: Yeah, that's a good idea. So, separate the desks please. 29

[students talking]

S: Can I take that pencil [TT]? 29

TT: Can you have - ?

S: The pencil?

TT: Yes. If you're going to need to borrow a pencil, come and get one now, thank 37
you.

S: ... a rubber?

TT: Oh well, I'll just put a rubber on each desk. Pencils are here, if you want one 37
borrow one, but it must come back. Right. When you get your booklet, do not do 29, 36
anything. Excuse me [7FS2], [7FS1], need you sitting now [7FS3] Right, now 29
29

C4: Student questionnaire responses.

Question	Year 7			Year 9		
1 How many of these tests have you done?	2 tests: 3 students					
	3 tests: 5 students			3 tests: 17 students		
	4 tests: 1 student					
	5 tests: 2 students					
	9 tests: 1 student					
	16 tests: 1 student					
	Yes	No	Unsure	Yes	No	Unsure
2 (a) Do you think these tests help you to learn better or allow yourself or your teachers to identify areas of your learning progress that need more attention?	12	1	0	13	2	2
2 (b) Do you or your teachers ever follow up on these?	15	2	0	0	17	9
3 Do you take these tests seriously?	11	3	0	16	1	0
4 Do you think your teachers take these tests seriously?	12	2	0	9	2	5
5 Any additional comments you would like to make about standardised/ benchmarking literacy tests or testing, and/or the ways they are used in your classroom. (Note: no Year 7 responses.)	'The tests want that bad but i dont like it when you have like 3 in a day it gets stress full when you try your best.' (Year 9)					
	'Can they be less boring.' (Year (9)					
	'give the teachers a card on how they went for their class to see where students are at.' (Year 9)					

Appendix D: List of Codes by Frequency

D1: Codes by frequency: Interview transcripts.

Codes	TT, Int. 1	TT, Int. 2	TT, Int. 3	T1, Int. 1	T2, Int. 1	P, Int. 1	TOTAL
1	5	4	9	7	7	0	32
2	2	2	4	2	2	2	14
3	6	0	8	7	10	11	42
4	3	0	4	4	10	11	32
5	1	0	2	5	1	0	9
6	1	1	1	3	0	0	6
7	1	0	2	1	5	6	15
8	19	3	4	0	4	3	33
9	20	4	18	13	12	13	80
10	6	1	6	2	0	0	15
11	6	11	21	23	7	3	71
12	1	0	1	0	1	0	3
13	0	0	0	0	10	0	10
14	1	0	1	2	1	0	5
15	21	6	9	0	0	0	36
16	2	0	0	0	0	4	6
17	20	4	7	2	2	1	36
18	19	0	1	6	2	7	35
19	17	5	3	2	0	4	31
20	0	0	3	13	8	5	29
21	5	0	0	4	2	0	11
22	2	1	3	2	5	0	13
23	1	14	9	6	2	2	34
24	0	0	1	0	1	6	8
25	1	0	0	4	13	6	24
26	0	0	6	0	3	0	9
27	0	0	12	2	10	10	34
28	0	4	15	7	9	7	42
29	0	0	5	9	6	0	20
30	0	1	6	5	2	0	14
31	1	0	7	8	3	12	31
32	1	0	4	11	2	0	18
33	0	0	4	0	5	5	14
34	1	2	4	6	6	0	19
35	0	0	10	12	8	2	32
36	0	0	1	0	1	0	2
37	0	0	0	0	0	5	5
38	0	2	4	8	0	5	19
39	0	0	0	0	1	3	4
40	0	0	0	0	0	4	4
41	0	3	0	9	0	0	12
42	3	0	0	13	3	0	19

43	0	1	3	0	7	0	11
44	0	0	4	0	4	0	8
TOTAL	166	69	202	200	175	137	947

D2: Codes by frequency: Testing session transcripts.

Codes by page: Testing session transcripts							
Codes	Test 1 (a, b)	Test 2	Test 3	Test 4	Test 5	Test 6	TOTAL
1	0	0	0	1	0	2	3
2	7	10	16	10	11	0	54
3	0	0	0	1	1	0	2
4	0	0	1	1	0	2	4
5	4	6	6	3	4	0	23
6	5	0	2	1	1	0	9
7	17	6	8	4	7	2	44
8	1	1	3	1	4	3	13
9	2	3	0	0	0	0	5
10	3	3	13	3	4	0	26
11	2	4	6	2	2	1	17
12	0	0	0	0	0	4	4
13	0	0	1	1	4	4	10
14	12	11	21	13	36	6	99
15	1	1	4	1	13	2	22
16	21	13	17	21	6	9	87
17	1	0	2	0	0	0	3
18	0	0	0	1	0	0	1
19	11	9	4	10	10	4	48
20	10	3	3	0	5	0	21
21	39	5	17	21	40	17	139
22	1	0	0	2	3	0	6
23	9	7	2	8	13	3	42
24	1	1	2	0	10	1	15
25	1	1	0	0	4	3	9
26	3	4	4	10	16	2	39
27	0	0	0	0	6	1	7
28	11	4	4	0	6	0	25
29	41	45	29	51	68	33	267
30	2	0	2	0	2	3	9
31	8	6	1	3	12	5	35
32	0	0	0	0	3	1	4
33	2	6	4	2	9	8	31
34	1	1	1	2	4	3	12
35	0	1	0	7	9	8	25
36	28	21	37	22	39	17	164
37	15	11	27	28	40	21	142
38	7	10	3	10	18	4	52
39	2	3	0	1	11	0	17
TOTAL	268	196	240	241	421	169	1535

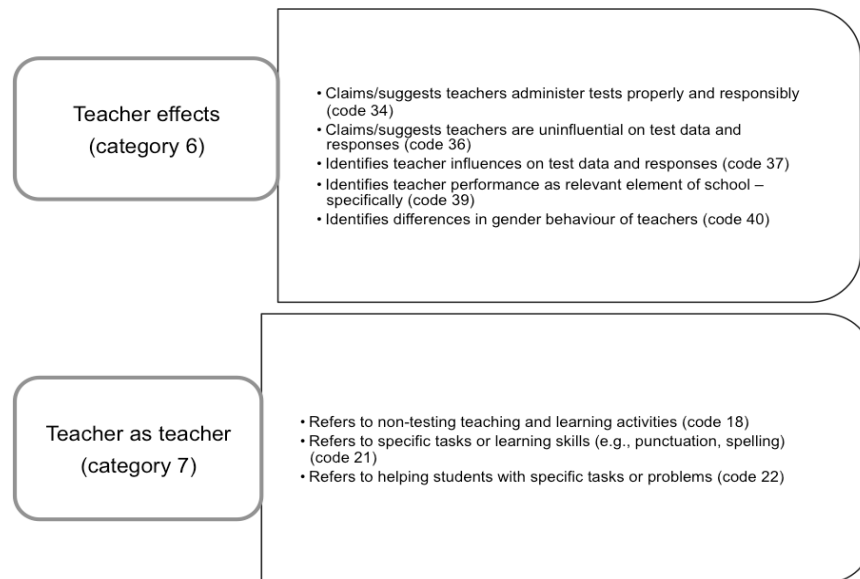


Figure 4-2. *Interview transcripts: Codes and categories.*

The seven categories constructed from the focused coding of the participant interview transcripts showed the broad groupings of participants' opinions and accounts of testing at their school site. Participants expressed various criticisms and critiques of standardised literacy tests and testing, regarding test documents, processes, requirements, relevance, appropriateness, uses, usefulness, value and validity. Participants also discussed issues not specific to standardised literacy testing, such as common teaching expectations and behaviours, presented opinions of their effects upon their students' progress (during testing and in a broader sense), and communicated judgements of their students according to student catchment and student academic and behavioural abilities.

Pragmatics of standardised literacy testing administration, and administrative uses (category 1)	<ul style="list-style-type: none"> • Over-identifies administration as feature/issue of tests (code 16) • Refers to the administrative uses of test data (code 17) • Reads aloud test instructions (code 36)
Teacher indecision as to the benefits of standardised literacy testing (category 2)	<ul style="list-style-type: none"> • Presents tests as potentially useful in student learning and the administration of learning topics (code 8) • Presents tests as positive incidents/tools but in a negative way (code 9) • Presents self as noncommittal, blasé or uncaring as to the tests, their uses and/or their administration (code 10) • Displays uncertainty as to the uses and administration of the tests (code 11)
Positive correlation between standardised literacy testing and student learning outcomes (category 3)	<ul style="list-style-type: none"> • Presents tests in a positive way (code 1) • Presents tests as useful/helpful for students and learning (code 3) • Presents tests as useful in the administration of learning topics for teachers (code 4) • Presents tests as beneficial, specifically to student learning outcomes (code 13) • Presents tests as an indicator of learning progress (code 15)
Standardised literacy testing as negative, problematic, lacking in educational merit (category 4)	<ul style="list-style-type: none"> • Presents tests in a negative way (code 2) • Presents tests as isolated incidents or tools (code 5) • Presents tests as not thoroughly utilised in student learning or the administration of learning topics (code 6) • Presents tests as problematically administered (code 7) • Presents or refers to the practice of tests as different to the perceived ideal (code 20)
Judgements on student communities and student catchment (category 5)	<ul style="list-style-type: none"> • Presents tests as beneficial for employment and post-school (code 12) • Refers to students as 'client-base', 'clientele' (code 18) • Categorises/judges schools on behavioural issues based on student catchment (code 22)
Judgements on student standards, learning levels (category 6)	<ul style="list-style-type: none"> • Presents tests as challenging and difficult (code 14) • Presents tests as revealing of student difficulties (code 24) • Refers to student skills (code 26)

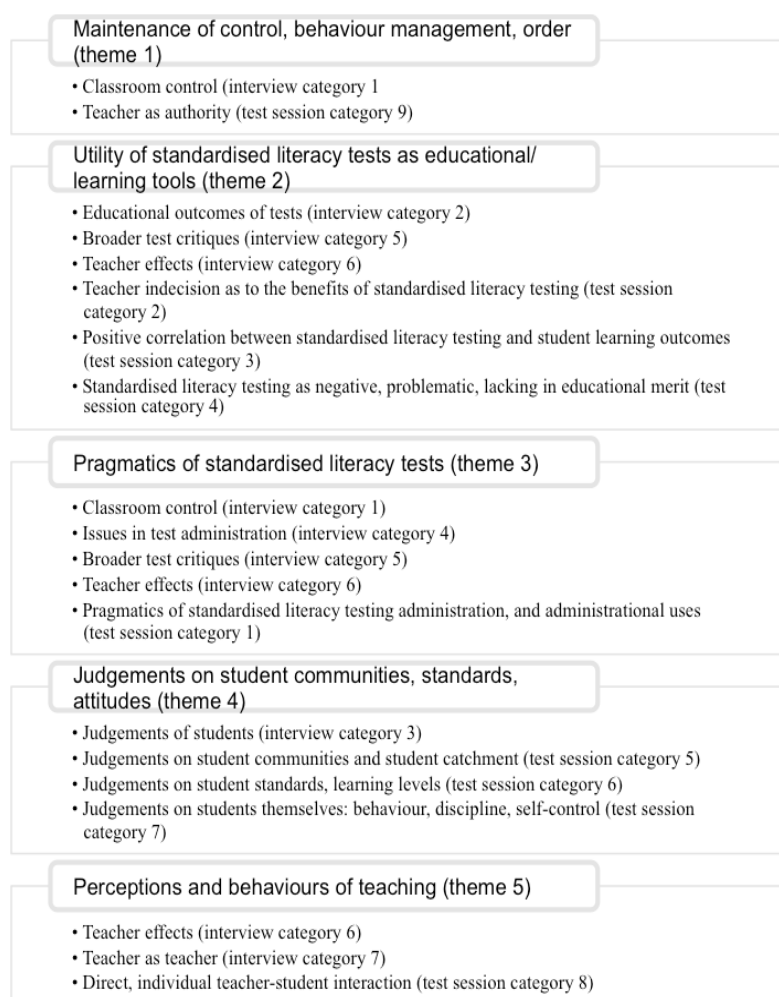


Figure 4-5. *Themes and categories.*

4.4.1 Theme 1: Maintenance of control, behaviour management, order.

The first theme, ‘Maintenance of control, behaviour management, order’, consists of only two categories, ‘Classroom control’ (interview category 1) and ‘Teacher as authority’ (test session category 9). The central theme connecting these codes and categories was the importance that participant teachers accorded classroom management. Within this theme, the data communicated that effective classroom management was highly important, in order to complete testing tasks and administer learning activities.

Appendix A: Approval Documentation

A1: Approvals: University of Tasmania Human Research Ethics Committee and Department of Education (Tasmania).

A1-a: University of Tasmania Human Research Ethics Committee approval letter.



MINIMAL RISK ETHICS APPLICATION APPROVAL

27 February 2006

Dr Claire Hiller
School of Education
Private Bag 66
Hobart

Ethics reference: H8730

'Standardised literacy testing: Issues, approaches, effects'

PhD candidate: Lauren Johnson

Dear Dr Hiller

Acting on a mandate from the Tasmania Social Sciences HREC, the Chair of the committee considered and approved the above project on 24 February 2006.

All committees operating under the Human Research Ethics Committee (Tasmania) Network are registered and required to comply with the *National Statement on the Ethical Conduct in Research Involving Humans 1999* (NHMRC guidelines).

Therefore, the Chief Investigator's responsibility is to ensure that:

- 1) All researchers listed on the application comply with HREC approved application.
- 2) Modifications to the application do not proceed until approval is obtained in writing from the HREC.
- 3) The confidentiality and anonymity of all research subjects is maintained at all times, except as required by law.
- 4) Clause 2.37 of the National Statement states:
An HREC shall, as a condition of approval of each protocol, require that researchers immediately report anything which might warrant review of ethical approval of the protocol, including:
 - a) *Serious or unexpected adverse effects on participants;*
 - b) *Proposed changes in the application; and*
 - c) *Unforeseen events that might affect continued ethical acceptability of the project.*

The report must be lodged within 24 hours of the event to the Ethics Executive Officer who will report to the Chairs.

A1-1 (continued)

- 5) All participants must be provided with the current Information Sheet and Consent form as approved by the Ethics Committee.
- 6) The Committee is notified if any investigators are added to, or cease involvement with, the project.
- 7) This study has approval for four years contingent upon annual review. An *Annual Report* is to be provided on the anniversary date of your approval. Your first report is due [12 months from 'Ethics Committee Approval' date]. You will be sent a courtesy reminder by email closer to this due date.
Clause 2.35 of the National Statement states:
As a minimum an HREC must require at regular periods, at least annually, reports from principal researchers on matters including:
 - a) Progress to date or outcome in case of completed research;
 - b) Maintenance and security of records;
 - c) Compliance with the approved protocol, and
 - d) Compliance with any conditions of approval.
- 8) A *Final Report* and a copy of the published material, either in full or abstract, must be provided at the end of project.

Yours sincerely



Ethics Executive Officer

A1-b: Department of Education (Tasmania) approval letter.

Department of Education
EDUCATIONAL PERFORMANCE SERVICES
2/99 Bathurst Street, Hobart
GPO Box 169, Hobart, TAS 7001 Australia



File: 650295

20 July 2007

Dr Claire Hiller
Senior Lecturer
School of Education
University of Tasmania
Private Bag 66
HOBART TAS 7000

Dear Dr Hiller

Standardised Literacy Testing: Issues, Approaches, Effects – Ms Lauren Johnson

I have been advised by the Educational Performance Research Committee that the above research adheres to the guidelines established and therefore have no objection to approving this study retrospectively. In order to update our records it would be appreciated if you could forward to us a brief outline of your progress to date.

We note that you sought permission from the Principal of [REDACTED] High School prior to your interviews with the English/Literacy teachers and that you conducted all interviews in an appropriate manner.

A copy of your final report should be forwarded to Patricia Lloyd, Educational Performance Services, Department of Education, GPO Box 169, Hobart 7001 at your earliest convenience and within six months of the completion of the research phase in Department of Education schools.

Yours sincerely

Manager
(Educational Performance Services)
Cc Ms Lauren Johnson

Appendix C: Data Extracts

C1: Extract from researcher's field notes.

P, Initial Pre-Interview

Thursday, 6 July 2006, 3 pm – 4 pm

- Data from the tests (OER) at [school] formed the ELs norms for the literacy and numeracy standards and progressions.
- The results come out so late in the year that they can only be used mainly for programme planning than for working with students individually about problem areas.
- The results report on reading, writing, overall literacy, and overall numeracy. They used to give more detail, breaking down into strands (those in the Statement and Profiles and other English curricula and syllabi), which was more helpful.
- Indigenous Education Unit: looks at indigenous students performing/scoring below benchmarks.
- The ENI is the numerical value scores for schools: schools given numbers between 0 and 120, usually between 18 and 90. The higher the number, the lower the SES, demographic, and so on.
 - [...] Primary schools ([school's] feeder primary schools) usually get the 80s and 90s.
 - [school and similar schools] high schools usually get ENI of 66.
 - [nearby school] HS scores in the low 50s.
- School size – all comparisons (ask OER)
- 2005 Yr. 7 Monitoring Test Results:

C2: Coded extract from interview transcript.

C2-a: TT, Interview 1, p. 1

TT, Interview 1: 18'51"

Wednesday, 19 July 2006, 9.25-9.45 am

Yr. 9 English/Humanities

TT: Er, give me five minutes to have this conversation, and then I'll have a look at it.

R: Okay, do you mind just holding that?

TT: Holding it? Alright, no trouble. Um – are we recording? Yes. Uh, the short answer is that, every teacher does it differently. So that what I do might be completely different to what the person in the room next door is doing.

R: Yep.

TT: And that, to a large extent, literacy and numeracy tests are the sorts of things that appear in the middle of the year. I do no explicit preparation for literacy tests, um, we use them as a diagnostic tool once the tests have been completed, in that we might look and see that individual X has a particular weakness that we may not have noticed. Um, the other thing we often use them for is that the results are used to confirm or to make us question the ratings we are currently giving [the students]. Other teachers may use them for much deeper diagnostic activities, but in my case I tend to use – they tend to be treated as a bit of an aside, as the sort of thing that you go through the process, and make sure the tests are administered as best you can, and then, if there's something that really jumps out at you, when the results come back, then you will, uh, take some action, according to that information. But, in general terms, they are the sorts of things that we use more as a school-wide indicator of how we're progressing, rather than as an individual diagnostic activity. I think that pretty much sums it up!

R: [chuckles]

TT: The other thing you need to know is that we're dealing with a client-base here where work ethic isn't high, in a lot of the kids. What you're seeing at the moment is really good ... but you're also seeing some individuals with really, really uh ... a heterogeneous mix of individuals, uh, from the very brightest to those who are still at early primary school level – with their literacy – all in one room. Which means that it's particularly difficult to stand up the front of the

C2-b: T1, Interview 1, p. 1

T1, Interview 1

Wednesday, 15 November 2006

- T1: Yes, it seemed like that's the first time they'd had to do one of these, they were used to calling out in class 'what's the answer? What's the answer?' They want you to tell them the answer, and when you tell them you couldn't and they – it was quite distressing for some of them. And it was really quite difficult. Other kids, a few in the grade 7 class which I did – I think the grade 7s handled it better than the grade 9s, in fact. We had a couple of anxiety attacks from a very capable student, who – in the writing exam, one young man, who is a very capable student, very capable – almost to the stage where he couldn't write because of the pressure to 'come up with something, *now*'. But he wrote a page and a half in the end, and I read it afterwards, it was very good. I saw the results, the test results came through the other day, he got an appropriate standard for that, I think it was a 5 Lower or something: he did well. But the anxiousness, the anxiety, brought on by the fact that 'I have to perform, well, and to a timeframe'. They just found that really hard.
- R: Yeah. In the transcripts, one of the testing sessions that I was in, I can't remember which grade it was, but they were really struggling with the concept that they had to do it in this certain amount of time. That they had this enforced-creativity, hour-and-a-half sort of thing, and they were saying, 'what happens if I don't finish it?' And they were told, 'well, you have to finish it'. 'But what if I can't, what if I can't?!' And they did manage to finish it, but I don't think they were very impressed by having those sort of limits pushed on them, that they had to do it, in that time.
- T1: I think one of the things that got me about it, I'll just refer back to something I did in class. We did a writing exercise, a spontaneous exercise, I'd put four words on the board, a noun, an adjective, and perhaps another noun and another adjective, and they had to write for twenty minutes, using those four words. And the first question that I would get, and I've done that in each of my classes ... 'how much do you want?' 'What do you want, and then I'll try and supply it.' ... trying to

C2-c: T2, Interview 1, p. 1

T2, Interview 1

Wednesday, 15 November 2006

[brief unrelated conversation between T2 and R prior to interview conversation]

R: I know that the questionnaire's a bit long and very detailed, but anything that you're ... that's fine.

T2: *[the audio-recorder is small]*

R: I know. I've got a smaller one at home, but it's so small that it just kind of gets lost in my bag. I've got a little list of interview questions, and some of them overlap from the questionnaire. The first one is just asking about procedures that you might follow when you're doing the tests in the classroom. I know that some teachers follow the instructions perfectly, thinking that that's well what they're supposed to do or some teachers might go over some things a bit more.

T2: Yeah, it's actually – I mean, they give you quite a lot of things, especially for the writing questions, that you go through to actually prepare them, the problem being of course that when you have three different tests to do and the numeracy tests and they're all supposed to be done within the one week, you could actually spend a whole lesson on that preparation stuff but you – the reality is that you can't really do that. There are things that I do which I'm sure aren't part of the

procedure, like, uh, sometimes I would read the word out to a kid who isn't sure if it's the word they think it is – I don't think you're really supposed to do that. But actually I think I only did that with the PAT – with the vocabulary tests, so that's a different set of tests than [you've/we've] just been doing. Some – there was something about – if you followed the instructions ... reading/comprehension – one of them was really not organised, I think it was that one, that if you just did what it says, the kids would have been quite confused, I think. So I don't follow every single thing to the letter.

R: But when you're not following it to the letter, you're not doing that because um – you'd be more likely to do that because you know what the students need to be able to –

C2-d: P, Interview 1, p. 1

P, Interview 1

Wednesday, 15 November 2006

- P: You came at an interesting time. The results have just been released last week and I'm in the process of sorting them out.
- R: That's very good timing for me!
- P: Yeah, it is. So what we're doing is – although the only thing that I've distributed to staff so far is this document, I don't know if [TT]'s shown it to you or not. 24
- R: Oh yeah, he showed me that.
- P: Which is an interesting piece of documentation in that by some strange process they've somehow correlated the performance in the reading, writing and spelling tests, and in the single numeracy test, into some projection towards a performance standard and award, which is a bit of a smoke and mirrors test to me. 24
- R: Well, I asked [TT] how he thought that they –
- P: But funnily enough, some of them are pretty good! 31
- R: Yeah, well! He said that they're pretty close.
- P: Some are, yeah. And the ... at the end which is saying that – that ... there, about the progressions statement, and progressions, and standards, is rather interesting. 3, 4, 7, 9, 20, 25
So it just says that they may not have shown, if you like, competence towards it. And it's significant, I've just picked up a grade nine class at the start of this term and we're doing some extension work, and one of the students, whose first pieces of work I graded at 5 Middle, he's only been assessed at a 4 Upper, so in terms of those kids at the upper end of the scale, it's not bad, but it depends on what sort of day they're having the day of the test, because there are other results I've gotten which are below and above, so it's – OER, to their credit – oh, well, yeah, to their credit – do make the point this is one test. They do make the point. For example, these are the score and percentile tabs. So they're giving a whole stack of this sort of stuff, and all too frequently, I think. Do you want to have a look, do you want these? 24, 28, 31
- R: Yeah, that'd be good.

C3: Coded extract from testing session transcript.

C3-a: Year Nine, Writing/Composition, p. 1

Year Nine Writing/Composition Test (session 1 of 2)

Tuesday, 25 July 2006, 9.00-9.45am

52'07" recorded

- TT: ... You've been introduced? You're not sure.
- S: Yeah.
- TT: [9MS1], ... so, it's one of those special things ... okay, you need to ... now, 29
okay!
- S: He's just going out for a smoke!
- TT: That means I don't want to hear another sound – as of now you are under 21
exam conditions, which means that you do not speak – 29
- S: ...
- S: ...
- TT: You do not speak, you do not spend any time worrying about what other 21, 29
people are doing, you are worried only about what you are doing. Now, I am
aware that for some of you, this is going to be a task that you don't enjoy, but 2
everybody in Tasmania has to do this, you're not the only people. You have to 14, 19, 26
do it to the best of your ability. This is not a test that has any influence on the 10
mark you get at the end of the year, it is a test that helps us as a school identify 15, 17
where people are with their literacy – with their reading, writing, spelling –
and helps us with some ideas about how we can help you.
- S: I need another pen. 9
- TT: So it's a really important test from the point of view of helping you. Now I'm
going to give you some instructions, you need to listen really carefully to these 16, 21, 29
instructions, and from now on, there should not be one word. The other really
important thing is to remember that this test is one that will take you possibly
60 or 65 minutes. However, it's quite possible you'll finish before that. You 29, 34
listening [9MS2]?
- S: Yeah.
- TT: If you finish before that, as I said to you about ten minutes ago in tutor, your 16, 29
job is to make sure that you have something sensible that you can be doing.
For most of you, the time is now nine o'clock, we're finishing here at 9.35, 14, 19
and we're gonna come back to it period four, for most of you it's going to be a
situation where you will not get finished during this lesson.

C3-b: Year Seven, Reading, p. 1

- Year Seven Reading Test
Tuesday, 8 August 2006, 10.55-11.50am
Yr. 7 Literacy Test 2006: Part A, Reading/Comprehension
T3 and T4: Other teachers
- TT: ... who are functionally illiterate. Do you want me to say all that again [*for the tape*]? 2, 14, 19,
- R: Okay there you go. 21, 23, 26,
- TT: Right. So we've got three kids who are functionally illiterate, and they're gonna 35
take a minute before they realise they can't do it, and then we're going to have to
keep them quiet whilst the others get on with it.
- R: Do you want me to take anything? 2, 5, 6, 7,
- TT: No, I think I'm right. So, that's a problem, and the other problem is time. There 10
isn't quite enough of it. But we'll do our best. [7FS1], that needs to go for this 29
lesson, please. If I see it, I'll have it. [Right you people need to spread yourselves
out, don't want to see people sitting beside each other. 16, 21, 29
- [noise and talking]
- TT: [R], on my desk is a box of rubbers, would you mind getting about another four?
Please? They're just sitting to the left of the computer.
- S: ...
- TT: Yeah, that's a good idea. So, separate the desks please. 29
- [students talking]
- S: Can I take that pencil [TT]? 29
- TT: Can you have - ?
- S: The pencil?
- TT: Yes. If you're going to need to borrow a pencil, come and get one now, thank 37
you.
- S: ... a rubber?
- TT: Oh well, I'll just put a rubber on each desk. Pencils are here, if you want one 37
borrow one, but it must come back. Right. When you get your booklet, do not do 29, 36
anything. Excuse me [7FS2], [7FS1], need you sitting now [7FS3] Right, now 29
29